This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n ⁶:
C12Q 1/68, G01N 33/53, C12P 19/34,
C12N 5/10, 1/21, C07K 5/00, 14/00,
16/00, C07H 21/02, 21/04

A1

(11) International Publication Number:

WO 96/25519

(43) International Publication Date:

22 August 1996 (22.08.96)

(21) International Application Number:

PCT/US96/01938

(22) International Filing Date:

15 February 1996 (15.02.96)

(30) Pri rity Data:

08/390,878

17 February 1995 (17.02.95) US

. |

(71) Applicant: PATHOGENESIS CORPORATION [US/US]; Suite 150, 201 Elliott Avenue West, Seattle, WA 98119 (US).

(72) Inventors: STOVER, Charles, Kendall; 7640 81st Place S.E., Mercer Island, WA 98040 (US). MAHAIRAS, Gregory, G.; 3312 39th West, Seattle, WA 98199 (US).

(74) Agents: HUNTER, Tom et al.; Townsend and Townsend and Crew, Steuart Street Tower, One Market, San Francisco, CA 94105-1492 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

(54) Title: VIRULENCE-ATTENUATING GENETIC DELETIONS

(57) Abstract

The present invention provides specific genetic deletions that result in an avirulent phenotype of a mycobacterium. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
ΑU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	ſΤ	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic		of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
СН	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Larvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

10

15

20

25

30

VIRULENCE-ATTENUATING GENETIC DELETIONS

BACKGROUND OF THE INVENTION

Mycobacterium nuberculosis (MTB) infects over ten million people each year and kills over three million, making it the infectious agent causing the greatest mortality worldwide. In an effort to combat Mycobacterium nuberculosis, vaccination programs using a viable attenuated strain of Mycobacterium bovis called bacille Calmette-Guérin (BCG) have been established in more than 120 countries over the course of the last 5 decades. Although widely used and considered safe enough to administer to infants, the BCG vaccine is controversial for two principle reasons: 1) Efficacy for BCG vaccines against tuberculosis has varied from 0-85% in different clinical trials; and 2) Immunization with BCG sensitizes vaccinees to the tubercular antigens used in the tuberculin skin test, confounding attempts to discriminate between BCG immunization and TB infection. For these two reasons, especially the latter, BCG is not used in the United States where surveillance with the tuberculin test is preferred.

The original Pasteur BCG strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations which cannot revert. However, the mutations which arose during serial passage of the original BCG strain have never been identified. Moreover, recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli have also been unsuccessful again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence. The antigenicity of BCG and the characteristics leading to its avirulence are thus poorly understood.

SUMMARY OF THE INVENTION

The present invention provides specific genetic deletions that account for the avirulent phenotype of the bacille Calmette-Guérin (BCG) strain of *Mycobacterium bovis*. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

10

15

20

25

In a preferred embodiment, this invention provides for nucleic acid sequences that are markers for avirulent or virulent mycobacteria. The sequences uniquely characterize the presence or absence of deletions that result in an avirulent phenotype. More specifically the sequence are either deletion junction sequence or deletion sequences or subsequences within deletion junction sequences or deletion sequences. Thus, this invention provides for a marker for an avirulent mycobacterium comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement includes BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. In a particularly preferred embodiment, the marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis of Mycobacterium bovis. or alternatively, the marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The marker may be the full length BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3 or a subsequence within any of these regions. The marker may also include a nucleic acid having at least 80%, preferably 90%. more preferably 95%, and most preferably 98% percent sequence identity with BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3. The marker may also include a sequence selected from an open reading frame of a the deletion sequences BCGa1, BCGa2, BCGa3. Suitable open reading frames are indicated in Figures 4, 5, and 6.

The above described marker may be a probe. The probe may be labeled by a number of means including, but not limited to radioactive, fluorescent, enzymatic, and colorimetric labels.

In another embodiment, this invention provides for polypeptides encoded by a subsequence of the BCGA1, BCGA2, or BCGA3 deletions. In particular, the subsequence may be selected from an open reading frame (ORF) present in one of these deletion sequences. This invention also provides for monoclonal or polyclonal antibodies that

10

15

20

25

30

specifically bind polypeptides encoded by one or more subsequences of the BCGa1, BCGa2, or BCGa3 deletions.

In still another embodiment, this invention provides for a recombinant cell comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGala, BCGala

In still yet another embodiment, this invention provides a method of distinguishing between an attenuated and a virulent mycobacterium. The method involves detecting the presence or absence of a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGAla, BCGAlb, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGAlab, BCGA2ab, BCGA3ab, BCGA1b, BCGA2, or BCGA3. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be

a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The polypeptide may be visualized by a number of means well known to those of skill in the art including antibody hybridization such as direct or indirect binding of labeled antibody.

5

10

15

This invention additionally provides a method for determining whether an attenuated or a virulent Mycobacterium is present in a sample. This method involves providing a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGala, BCGalab, BCGalab, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3; and hybridizing the first nucleic acid to the biological sample. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium suberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The method may also include detecting the hybridized first nucleic acid. This may involve direct detection of a label or additionally involve an amplification step and subsequent detection of the amplified product.

25

30

20

Finally, this invention provides a method of producing an attenuated-virulence mycobacterium. This method involves deleting from the genomic DNA of a virulent mycobacterium a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGA1, BCGA2, and BCGA3. The first nucleic acid may be BCGA1, BCGA2, or BCGA3, or alternatively, it may be a promoter, other control element or an open reading frame from BCGA1, BCGA2, or BCGA3.

Definitions

5

10

15

20

25

30

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The phrase "specifically detect" as used herein refers to the process of determining that a particular subsequence is present in a DNA sample. A DNA sequence may be specifically detected through a number of means known to those of skill in the art. These would include, but are not limited to amplification of the particular target sequence through polymerase chain reaction or ligase chain reaction, hybridization of the sequence to a labeled probe, and binding by labelled ligands or monoclonal antibodies. For a discussion of various means of detection of specific nucleic acid sequences see Perbal, B. A Practical Guide to Molecular Cloning, 2nd Ed. John Wiley & Sons, N.Y. (1988) which is incorporated herein by reference.

The phrase "select subsequence" is used herein to refer to a particular DNA subsequence that is of interest. It is often a predetermined or known sequence of nucleic acid bases. A select subsequence is typically chosen because of a unique sequence identity. Typically a select subsequence is targeted for DNA amplification and often is useful as a specific marker for the presence of a particular gene or a deletion of a particular nucleic acid sequence.

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides. Oligonucleotides may include, but are not limited to, primers, probes, nucleic acid fragments to be detected, and nucleic acid controls. Oligonucleotides include naturally occurring nucleotides, chemically modified naturally occurring nucleotides and synthetic nucleotides. The exact size of an oligonucleotide depends on many factors and the ultimate function or use of the oligonucleotide.

The term "primer" refers to an oligonucleotide, whether natural or synthetic, capable of acting as a point of initiation of DNA synthesis under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization (i.e., DNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. A primer is preferably a single-stranded oligodeoxyribonucleotide.

The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 25 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template.

The phrase "PCR primers competent to amplify" as used herein refers to a pair of PCR primers whose sequences are complementary to DNA subsequences immediately flanking the DNA subsequence (target sequence) which it is desired to amplify. The primers are chosen to bind specifically those particular flanking subsequences and no other sequences present in the sample. The PCR primers are thus preferably chosen to amplify the unique target sequence and no other. Alternatively, the PCR primers may be selected to bind to sequences other than the target sequence where the amplification products can be subsequently distinguished (e.g. where the desired amplified sequence is different in size than other amplified sequences).

15

20

25

30

10

5

"Amplifying" or "amplification", which typically refer to an "exponential" increase in target nucleic acid, are used herein to describe both linear and exponential increases in the number of a select target sequence of nucleic acid.

The term "antisense orientation" refers to the orientation of nucleic acid sequence from a structural gene that is inserted in an expression cassette in an inverted manner with respect to its naturally occurring orientation. When the sequence is double stranded, the strand that is the template strand in the naturally occurring orientation becomes the coding strand, and vice versa.

The term "deletion" refers to a region of a nucleic acid which is not present in an organism, but which is present in another related organism. In the context of mycobacteria, a deletion refers, e.g., to a region of nucleic acid which is not present in one strain of mycobacteria, but which is present in another related strain. For instance, an avirulent mycobacterial strain can have a deletion in its genome relative to the genome of a related virulent mycobacterial strain.

The term "deletion junction" refers to the region of a nucleic acid spanning the insertion point of a deletion. Thus, where a region of a nucleic acid sequence is deleted (i.e. a deletion is present), the deletion junction spans the nucleotides that are immediately adjacent to the deletion. Conversely, where a region of a nucleic acid sequence is not

deleted (i.e. the deletion is absent), two deletion junctions are present, each spanning respectively one end of the deletion sequence and its flanking sequence.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Figures 1, 2, or 3, or may comprise a complete cDNA or gene sequence.

10

15

5

Generally, a reference sequence is at least 10 nucleotides in length, frequently at least 20 to 25 nucleotides in length, and often at least 50 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

20

25

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch J. Mol. Biol. 48: 443 (1970); by the search for similarity method of Pearson and Lipman Proc. Natl. Acad. Sci. (USA) 85: 2444 (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected.

30

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned

10

15

20

25

30

sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The terms "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. The isolated nucleic acid probes of this invention do not contain materials normally associated with their in situ environment, in particular nuclear, cytosolic or membrane associated proteins or nucleic acids other than those nucleic acids intended to comprise the nucleic acid probe itself.

The term "marker" refers to a characteristic which distinguishes one class of cells or compositions from a second class of cells or compositions. For instance, the deletions and deletion junctions described herein can be used to distinguish between strains (e.g., virulent and avirulent strains) of mycobacteria. While markers are indicators of associated features or properties, as used herein, markers may also be used for purposes other than indicating the associated feature or property. Thus, for example, a nucleic acid marker of virulence identifies a particular nucleic acid which may be used in a variety of contexts other than simply indicating virulence.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "peptide" or "polypeptide" refers to an amino acid polymer which is encoded by a nucleic acid. The peptide or polypeptide may include naturally occurring or modified amino acids.

٠. ﴿

43

1

5

10

15

20

25

30

The terms "probe" or "nucleic acid probe" refer to a molecule that binds to a specific sequence or subsequence of a nucleic acid. A probe is preferably a nucleic acid which binds through complementary base pairing to the full sequence or to a subsequence of a target nucleic acid. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarily with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labelled such with, e.g., biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

The term "labeled nucleic acid probe" refers to a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen "bonds" to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference to a cell indicates that the cell replicates or expresses a nucleic acid, or expresses a peptide or protein encoded by DNA whose origin is exogenous to the cell. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes found in the native form of the cell wherein the genes are re-introduced into the cell by artificial means.

The term "sample" refers to a material with which bacteria may be associated. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. It will be recognized that the term "sample" also includes supernatant from eukaryotic cell cultures (which may contain free bacteria), cells from cell or tissue culture, and other media in which it may be desirable to detect mycobacteria (e.g., food and water).

The term "subsequence" in the context of a particular nucleic acid sequence refers to a region of the nucleic acid equal to or smaller than the specified nucleic acid.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least 90% sequence identity to a reference sequence, or preferably 95%, or more preferably 98% sequence identity to the

10

15

20

25

30

reference sequence, over a comparison window of at least about 10 to about 100 nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptides which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

The term "uninterrupted reading frame" or "open reading frame" refers to a DNA sequence (e.g., cDNA) lacking a stop codon or other intervening, untranslated sequence. An intact open reading frame refers to a full length uninterrupted reading frame or minor variations thereof.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal at a rate that is detrimental to the cell or animal within its host range. More particularly virulent mycobacteria persist longer in a host than avirulent mycobacteria. Virulent mycobacteria are typically disease producing and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism. Typically, infection of guinea pigs is used as an assay for mycobacterial virulence. In contrast, the term "avirulent" refers to a bacterium or strain of bacteria that either does not replicate within a host cell or animal within its host range, or replicates at a rate that is not significantly detrimental to the cell or animal.

The term BCG-like avirulence, as used herein refers to an attenuated virulence brought about by one of the deletions of the present invention.

10

15

20

25

30

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete sequence listing of the BCG deletion region 1 including flanking sequences. The deletion, designated BCGa1, is located between nucleotide 2327 and nucleotide 11126.

Figure 2 shows the complete sequence listing of the BCG deletion region 2 including flanking sequences. The deletion, designated BCGA2, is located between nucleotide 3382 and nucleotide 14071.

Figure 3 shows the complete sequence listing of the BCG deletion region 3 including flanking sequences. The deletion, designated BCGA3, is located between nucleotide 1406 and nucleotide 10673. "N" represents "A", "C", "G", or "T".

Figure 4 shows a map of the deletion sequence BCGa1. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown.

Figure 5 shows a map of the deletion sequence BCGa2. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozomal binding sites and homologies to the predicted encoded proteins are shown.

Figure 6 shows a map of the deletion sequence BCGA3. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown. The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, remains to be determined. Therefore, the base pair coordinates given in the region 3 map 3' to the 9kb marker are approximations. The precise sequence determination of this region is likely to effect the length of open reading frames 3H and 3L.

Figure 7 illustrates the deletion junction regions of BCGa1, BCGa2, and BCGa3. The "terminal" deletion junction regions formed by the flanking sequences and the terminal regions of the deletion sequences are identified as BCGa1a, BCGa1b, BCGa2a, BCGa2b, and BCGa3a, and BCGa3b. When the deletion is present (the deletion sequences

10

15

20

25

30

are missing) the respective "a" and "b" sequences will be juxtaposed, thereby forming deletion "spanning" junction sequences designated BCGalab, BCGa2ab, and BCGa3ab, respectively.

Figure 8 shows EcoRI and BamHI restricted chromosomal DNAs from Mycobacterium bovis, BCG Connaught, and Mycobacterium tuberculosis strains H37Ra, H37Rv, and Erdman probed with ³²P labeled BCG subtracted probe.

DETAILED DESCRIPTION

This invention reflects the discovery of genetic deletions in mycobacteria that result in an avirulent genotype such as is exhibited by the bacille Calmette-Guérin (BCG) mycobacterium. The original Pasteur bacille Calmette-Guérin (BCG) strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations that cannot revert. The mutations that arose during serial passage of the original BCG strain were not previously known. Recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli were unsuccessful, again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence.

The genetic deletions leading to the avirulent phenotype of BCG were identified by genomic subtractions between Connaught strain of BCG and MBV/MTB. The subtracted probe resulting from the genomic subtraction between BCG and the H37 Rv strain of M. tuberculosis was subsequently used to identify and clone three regions from a cosmid library of Mycobacterium bovis genomic DNA. Southern blot mapping and DNA sequence comparisons between BCG and M. bovis showed that three regions, designated regions 1-3, contained DNA segments of approximately 9 kb, 11 kb and 9 kb respectively, which are deleted in the Connaught strain of BCG. Precise deletion junctions were identified for each region by comparisons of BCG and corresponding virulent MBV sequences. The respective deletions, designated BCGA1, BCGA2 and BCGA3 are illustrated in Figures 1-3.

One of skill in the art will appreciate that the deletions encompassed by BCGa1, BCGa2 and BCGa3 may be utilized in a variety of contexts. For example, the deletions may be utilized to distinguish between avirulent and virulent strains of

10

15

20

25

30

mycobacteria thereby providing early detection of patients at risk for tuberculosis. This is of particular importance where mycobacteria are identified in a sample from a patient that has been previously vaccinated with BCG. In this context it may be critical to determine whether mycobacteria identified in a biological sample from such a patient are pathogenic.

In another embodiment, the preparation of mycobacteria containing the deletions of the present invention may provide superior vaccines to BCG which has long been known to have marginal efficacy. Thus, for example, a *Mycobacterium tuberculosis* may contain a full BCGal deletion or a smaller deletion within BCGal (e.g. one or more open reading frames) rendering it avirulent. An avirulent MTB will provide a more efficient vaccine because it is antigenically more similar to MTB than is BCG. Moreover, an MTB rendered avirulent by the production of smaller deletions within the deletion regions identified in this invention will present more antigenic determinants.

Since the loss of virulence is due to the loss of gene products expressed by the nucleic acid sequences comprising the deletion regions, the BCGA1, BCGA2 and BCGA3 deletion sequences and proteins encoded within these deletion sequences provide suitable targets for drug screening. Thus, the use of deleted sequences as targets to screen for drugs that inhibit or interfere with transcription, translation, or post-translational processing of proteins encoded by the deletion sequences, or with the deletion encoded polypeptides themselves, provides an assay for anti-mycobacterial agents. In particular, the use of reporter genes such as firefly luciferase (FFlux), \(\beta\)-galactosidase (BGal), and the like, under the control of promoters present in the deletion sequence provide a rapid assay for drugs regulating activity originating in this region. Conversely, since the protein products of the deletion sequences are presumably expressed in virulent mycobacterial species, proteins expressed by deletion sequences may make good antigens for antimycobacterial vaccines.

Finally, as the viability of BCG demonstrates, deletion regions BCGa1, BCGa2 and BCGa3 are not required for mycobacterial growth and reproduction. Thus, these deletion regions provide good insertion points for the expression of heterologous DNA. The heterologous DNA sequences may be under the control of endogenous inducible or constitutive promoters typically found in the deletion sequences, or alternatively, they may be under the control of introduced promoters, either constitutive or inducible, exogenous to mycobacteria.

10

15

20

25

30

I. Detection of Deletions

As indicated above, the deletions identified in the present invention provide useful markers for the identification of an avirulent (or conversely a virulent) mycobacterial phenotype. Specifically, determination of avirulence simply requires the detection of the presence or absence of the deletion (either BCGA1, BCGA2, or BCGA3, or deletions within these regions). Where the deletion is present in the bacterial DNA, the bacterium expresses a BCG-like avirulent phenotype. Conversely, where the deletion is absent in the bacterial DNA, the bacterium does not express a BCG-like avirulence. While this may indicate that the bacterium is virulent, one of skill will appreciate that the bacterium may still be avirulent due to the presence of other mutations or deletions. Nevertheless, screening for the presence of the deletion provides a means of detecting a BCG-like avirulent mycobacterium.

Means of detecting deletions are well known to those of skill in the art. Generally, the deletions may be detected either by detecting the presence or absence of deletion junctions, or, alternatively, by detecting the presence or absence of the sequences contained within the deletion (deletion sequences). Where a nucleic acid sequence is deleted (i.e., a deletion is present), the sequences that previously flanked the deleted sequence are juxtaposed, thereby forming a new deletion junction that spans the deletion. Detection of the presence of such a "spanning" deletion junction indicates the presence of the deletion and thus the avirulent phenotype.

Conversely, where the nucleic acid sequence is not deleted (the deletion is not present) the spanning junction sequence will be absent (See, e.g. Figure 7). The "terminal" deletion junction sequences flanking each endpoint of the deletion region are present and detection of these terminal deletion junctions indicates the absence of a deletion. Spanning deletion junction regions and terminal deletion junctions suitable for detecting the deletions of the present invention are illustrated in Figure 7 and in Table 1.

Table 1. Nucleic acid sequences comprising deletion junctions. The symbol "|" indicates the insertion point of the deletion sequence. Deletion sequence bases are represented in lower case letters.

Junction	Nucleotide Sequence	Seq.
BCGAla	CTGGTCGACGATTGGCACAT gcagccgtgggtgccgccgg	1

BCG△lb	gigicticateggeticcae CCAGCCGCCCGGATCCAGCA	2
BCG ₄ 2a	CAACTCCACGGCGACCACCC gcgcccccgctcgcactaga	3
BCG△2b	gcccacccggtcgagcaccc CGATGATCTTCTGTTTGACC	4
BCG ₄ 3a	CACCTCGACCACGGCCAACC gtggacctgtgagatacact	5
BCG ₄ 3b	tcagcagtccacggccaacc CCGCACCAACACCTTCCACC	6
BCG△lab	CTGGTCGACGATTGGCACAT CCAGCCGCCCGGATCCAGCA	7
BCG _△ 2ab	CAACTCCACGGCGACCACCC CGATGATCTTCTGTTTGACC	8
BCG∆3ab	CACCTCGACCACGGCCAACC CCGCACCAACACCTTCCACC	9

15

5

Where a deletion is detected by determining the presence or absence of sequences contained within the deletion (deletion sequences), the absence of deletion sequences indicates the presence of a deletion and thus an avirulent phenotype. Conversely, the presence of deletion sequences indicates the absence of a deletion. Deletion sequences that provide suitable targets for detecting the deletions of the present invention are provided in Figures 1, 2 and 3.

43. **

A) Isolation of DNA for Detection of Mycobacterium Genomic Deletions

20

In a preferred embodiment, DNA is obtained from mycobacteria. As used herein, the term "mycobacteria" refers to any bacteria of the family Mycobacteriaceae (order Actinomycetales) and includes, but is not limited to, Mycobacterium tuberculosis, Mycobacterium avium complex, Mycobacterium kansasii, Mycobacterium scrofulaceum, Mycobacterium bovis and Mycobacterium leprae. These species and groups and others are described in Baron, S., ed. Medical Microbiology, 3rd Ed. (1991) Churchill Livingstone, New York, which is incorporated herein by reference.

25

The identification of deletions using a DNA marker requires that the DNA sequence be accessible to the particular probes used or to the components of the amplification system if the DNA sequence is to be amplified. In general, this accessibility is ensured by isolating the nucleic acids from the sample.

30

A variety of techniques for extracting nucleic acids from biological samples are known in the art. For example, see those described by Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New

York, (1985), by Han, et al. Biochemistry, 26: 1617-1625 (1987) and by Du, et al. Bio/Technology, 10: 176-181 (1992), which are incorporated herein by reference.

Alternatively, if the sample is readily disruptable, the nucleic acid need not be purified prior to amplification by the PCR technique, *i.e.*, if the sample is comprised of cells, particularly peripheral blood lymphocytes or monocytes, lysis and dispersion of the intracellular components may be accomplished merely by suspending the cells in hypotonic buffer or boiling them in a low concentration of alkali (*i.e.* 10 mM NaOH).

In a preferred embodiment, DNA is extracted from mycobacteria as described in Example 1.

10

15

20

25

30

5

B) Detection of Deletions Using Hybridization Probes

In one embodiment the avirulence deletions are detected by contacting DNA obtained from the mycobacterium with a probe that specifically binds an entire deletion junction region or a subsequence of that region and does not specifically bind to any other DNA sequences in the sample. Alternatively, a probe that specifically binds the entire deleted region or subsequence of that region and does not specifically bind to any other sequences in the sample is also suitable. While such probes may be proteins, oligonucleotide probes are preferred. Typically, the sequence of the oligonucleotide probe is chosen to be complementary to a select subsequence unique to the deletion junction or the deletion sequence, whose presence or absence is to be detected. Under stringent conditions the probe will hybridize with the select subsequence forming a stable duplex.

The probe is typically labeled. Detection of the label in association with the target DNA indicates either the presence or absence of the deletion. The probe may be used to detect the deletion junction or deletion sequences directly in a DNA sample without amplification of the deletion subsequences. In one embodiment, unamplified DNA sequences are probed using a Southern blot. The DNA of the sample is immobilized, on a solid substrate, typically a nitrocellulose filter or a nylon membrane. The substrate-bound DNA is then hybridized with the labeled probe under stringent conditions and non-specifically hybridized probe is washed away. Labeled probe detected in association with the immobilized mycobacterial sequences (e.g. bound to the substrate) indicates the presence of deletion sequences (e.g. BCGa1, BCGa2, or BCGa3) and therefore the absence of the deletion. Means for detecting specific DNA sequences are well known to those of skill in

10

15

20

25

30

the art. Protocols for Southern blots as well as other detection methods are provided in Maniatis, et al. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY (1982), which is incorporated herein by reference.

In another embodiment, the mycobacterial DNA subsequences are themselves labeled. They are then hybridized, under stringent conditions, with a probe immobilized on a solid substrate. Detection of the label in association with the immobilized probe indicates the presence or absence of the deletion.

In a preferred embodiment, the deletion junction sequences or subsequences or the deletion sequences or subsequences may be amplified by a variety of DNA amplification techniques (for example via cloning, polymerase chain reaction, ligase chain reaction, transcription amplification, etc.) prior to detection using a probe. Because the copy number of mycobacterial sequences bearing the virulence-attenuating deletions is low, the use of unamplified mycobacterial DNA results in an assay of low sensitivity. Amplification of mycobacterial DNA increases sensitivity of the assay by providing more copies of possible target subsequences. In addition, by using labeled primers in the amplification process, the mycobacterial DNA sequences are labeled as they are amplified.

C) Selection of Probes for Detection of the Deletion Junction Sequences or the Deletion Sequences

Full length sequences are provided for the deletions BCGa1, BCGa2, and BCGa3 in Figures 1, 2 and 3 respectively. Using these sequence listings, one of skill in the art may easily determine appropriate probes or primers for the detection of the presence or absence of the deletion junctions or the deletion sequences. Generally speaking, a probe will be selected that hybridizes to the target junction sequences or deletion sequences, but not to other mycobacterial nucleic acid sequences under stringent conditions. The design of hybridization probes is well known in the art. See, for example, Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989), which is incorporated herein by reference.

In a preferred embodiment, the probe is an oligonucleotide sequence complementary to a subsequence comprising a deletion junction (e.g. BCGala, BCGalb, BCGala, BCGala, BCGalab, BCGala

10

15

20

25

30

sequence complementary to a subsequence of a deletion sequence (e.g. BCGa1, BCGa2, and BCGa3). The probe preferably has destabilizing mismatches with subsequences from other regions of the mycobacterial genome.

The exact length of the probe depends on many factors including the length of conserved regions around the deletions, the degree of sequence specificity desired, and the amount of internal complementarity within the probe. Such probes are preferably 17 to 25 bases in length. One of skill will recognize that longer probes specifically hybridize at higher temperatures. Generally, stringent conditions are selected to be about 5°C to 20°C, more preferably about 10°C, lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from an avirulent mycobacterium such as BCG, but not to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV. Alternatively, Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from an avirulent mycobacterium such as BCG.

Oligonucleotide probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang et al. Meth. Enzymol, 68: 90-99 (1979); the phosphodiester method of Brown et al., Meth. Enzymol. 68:109-151 (1979); the diethylphosphoramidite method of Beaucage et al., Tetra. Lett., 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

Probe detectability may be increased by the attachment of a label. As used herein, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g. DynabeadsTM), fluorescent dyes (e.g., fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads.

Methods for attaching labels to probes, primers, and antibodies are well known to those of skill in the art. For example, the probe can be labeled at the 5'-end with ^{32}P by incubating the probe with ^{32}P -ATP and polynucleotide kinase (see Perbal, A

Practical Guide to Molecular Cloning, 2nd ed. John Wiley, N.Y. (1988)). Other labels may be joined to the probe directly or through linkers. They may be located at the ends of the probe or internally. Methods of attaching labels may be found in Connell, et al., Bio/Techniques 5: 342 (1987), U.S. Patent Nos. 4,914,210, 4,391,904 and 4,962,029, which are incorporated herein by reference. In addition, kits for labelling oligonucleotides are widely available. See, for example, Boehringer Mannheim Biochemicals (Indianapolis, IN) for "Genius" labeling kits based on dioxigenin technology and Clonetech (South San Francisco, CA) for a variety of direct and indirect oligonucleotide labeling reagents.

10

15

20

25

30

5

D) Detection of Deletions Conferring Avirulence Through Amplification of Unique Subsequences

Deletions are particularly amenable to detection without the use of a hybridization probe. In a preferred embodiment, subsequences are amplified that include a deletion junction. The amplified deletion junction may be a "spanning" deletion junction in which case where the deletion is present (i.e. the deletion sequences are absent), the amplification product is a specific DNA incorporating the deletion junction sequence spanning the deletion (e.g. incorporating flanking sequences from both sides of the deleted sequence). Where the deletion is absent (i.e. deletion sequences are present) and primers are selected so that there are no priming sites within the deletion sequences, amplification is non-existent or alternatively provides a complex mixture of non-specifically amplified fragments. Alternatively, amplification primers may be selected that specifically hybridize to deletion sequences, as long as they are selected to amplify sequences that are distinguishable from the sequence amplified when the deletion is present.

Alternatively, the amplification product may be subsequence of a "terminal" deletion junction in which case absence of the deletion (i.e. the deletion sequences are present) will result in the amplification of the specifically targeted nucleic acid. Conversely, where the deletions are present (i.e. the deletion sequences are absent) there will be no specific amplification of a terminal deletion junction.

Amplification products may be separated by size for characterization. Size separation may be accomplished by a variety of means known to those of skill in the art.

10

These methods include, but are not limited to electrophoresis, density gradient centrifugation, liquid chromatography, and capillary electrophoresis. In a preferred embodiment, the fragments are separated by agarose gel electrophoresis. The bands are then stained with a marker to visualize them such as ethidium bromide and the gel is visualized, e.g., using ultraviolet light.

As described above, an agarose gel typically shows I band if the deletion is present, reflecting amplification of the deletion-spanning sequence. Where the deletion is absent, amplification results in either no bands, where there are no sequences within the deletion to which the amplification primers may hybridize, or a smear where there is non-specific amplification, or a series of discrete bands distinguishable from the band representing the deletion-spanning sequence where primers are chosen that hybridize to deletion sequences.

E) Selection of Primers for Amplification of Avirulence Deletions

15 Amplification of deletion junction sequences or subsequences or deletion sequences or subsequences may be accomplished by methods well known in the art, which include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990), which is incorporated herein by reference), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and 20 Barringer, et al., Gene, 89: 117 (1990), which are incorporated herein by reference), transcription amplification (see Kwoh, et al., Proc. Natl. Acad. Sci. (U.S.A.), 86: 1173 (1989) which is incorporated herein by reference), and self-sustained sequence replication (see Guatelli, et al., Proc. Nat. Acad. Sci. (U.S.A.), 87: 1874 (1990) which is incorporated herein by reference), each of which provides sufficient amplification so that 25 the target sequence can be detected by nucleic acid hybridization to a probe or by electrophoretic separation. Alternatively, methods that amplify the hybridization probe to detectable levels can be used, such as $Q\beta$ -replicase amplification. See, for example, Kramer, et al. Nature, 339: 401 (1989), Lizardi, et al. Bio/Technology, 6: 1197 (1988), and Lomell, et al., Clin. Chem. 35: 1826 (1989) which are incorporated herein by 30 reference.

In a preferred embodiment, amplification is by polymerase chain reaction using a pair of primers that flank and thereby amplify a selected deletion junction subsequence. Selection of primers is readily apparent to one of skill in the art using the sequence listings of the present invention. For example, a pair of PCR primers 5'-TCGACGATTGGCACAT-3' (T_m =55°C) and 5'-TCCCTCCCTGTATTTGTAT-3' (T_m =56°C) will amplify a 469 base pair sequence including the BCGala deletion junction, while 5'-CGTTCTTCGGAGGTTTC-3' (T_m =56°C) and 5'-GGCGGCTGGGTGGA-3' (T_m =60°C) will amplify a 471 base pair sequence including the BCGalb deletion junction.

10

15

20

25

5

F) Detection of Deletions through Detection of Expression Products of Deletion Sequences

In addition to the detection of deletions by the detection of either the deletion junction sequences or the deletion sequences, one may detect the absence of the deletion by detecting the expression products of the deletion sequences. Thus, for example, where the deletion sequences express a protein, the presence of that protein indicates the absence of the deletion and thus is indicative of a virulent (non BCG-like) phenotype. Such proteins are referred to herein as "deletion polypeptides".

Means of determining proteins expressed by particular nucleic acid sequences are well known to those of skill in the art. Typically this involves determining the longest open reading frame. This may be aided by the identification of initiation sites (e.g. ribozome binding sites). The protein encoded by the largest open reading frame is determined using codon preferences for the specific organism from which the nucleic acid is obtained. The polypeptide sequence listing may then be compared against a sequence database, e.g. GenBank, to determine other sequences sharing substantial sequence identity with the calculated sequence. The expression of the protein may be verified by isolating and then sequencing proteins having the predicted length and charge characteristics.

30 method then de

Once deletion polypeptides are identified they may be detected by routine methods well known to those of skill in the art. Typically this involves isolating and then detecting the polypeptide. The polypeptide may be isolated by a number of means well known to those of skill in the art. This includes typical methods of protein

purification such as high performance liquid chromatography (HPLC), electrophoresis, capillary electrophoresis, hyperdiffusion chromatography, thin layer chromatography, and the like. Methods of purifying and detecting proteins are well known to those of skill in the art (see, e.g., Methods in Enzymology Vol. 182: Guide to Protein Purification, M. Deutscher, ed. Vol. 182 (1990), which is incorporated herein by reference).

Alternatively, deletion polypeptides sequences may be detected using immunoassays utilizing antibodies specific for the deletion polypeptides. The production of such antibodies and their use in immunoassays is detailed below.

G) Antibodies to Deletion Polypeptides

Antibodies can be raised to the polypeptides encoded by the nucleic acids corresponding to the open reading frames present in the deletion regions of the present invention (deletion polypeptides). As used herein "antibodies" include immunoglobulin or a population of immunoglobins which specifically bind to an antigen. Thus an antibody may be monoclonal or polyclonal including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies can be raised to these polypeptides in either their native configurations or in non-native configurations. Anti-idiotypic antibodies may also be used.

20

25

30

15

5

10

1) Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with deletion polypeptides. Recombinant polypeptides are the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring polypeptides may also be used either in pure or impure form. Synthetic peptides made using sequences described herein may also used as immunogens for the production of antibodies.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide is injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

200.4

5

10

15

20

25

30

Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified deletion polypeptide is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY, which are incorporated herein by reference.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256: 495-497, which discusses one method of generating monoclonal antibodies.

Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells (See, Kohler and Milstein (1976) Eur. J. Immunol. 6: 511-519, incorporated herein by reference). The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro.

Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al. (1989) Science 246: 1275-1281. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B

10

15

20

25

30

cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

24

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, Huse et al. Science 246: 1275-1281 (1989); and Ward, et al. Nature 341: 544-546 (1989). The polypeptides and antibodies of the present invention are used with or without modification, including chimeric antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen et al. Proc. Nat'l Acad. Sci. USA 86: 10029-10033 (1989).

Antibodies, including binding fragments and single chain versions, against predetermined fragments of deletion polypeptides can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptides, or screened for agonistic or antagonistic activity, e.g., mediated through a receptor. These monoclonal antibodies will usually bind with at least a K_D of about 1 mM, more usually at least about 300 μ M, and most preferably at least about 0.1 μ M or better.

The antibodies of this invention can also be used for affinity chromatography in isolating deletion polypeptides. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a bacterial lysate, or recombinant cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified deletion polypeptides are released.

10

15

20

25

30

The antibodies can be used to screen expression libraries for particular expression products. Usually the antibodies in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

In a preferred embodiment, antibodies to deletion polypeptides are used for the identification of cell populations expressing the polypeptides. By assaying the expression products of cells expressing the polypeptides it is possible to diagnose bacterial infections.

Antibodies raised against each polypeptide are useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to the presence of the respective antigens.

2) Immunoassays

A particular deletion polypeptide can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) 1991 Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in Maggio (ed.) (1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is incorporated herein by reference. See also Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed.) (1988) Non-isotopic Immunoassays Plenum Press, NY.

Immunoassays for measurement of deletion polypeptides can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be, e.g., competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with a deletion polypeptide produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

WO 96/25519 PCT/US96/01938

26

In a competitive binding immunoassay, the deletion polypeptide present in the sample competes with labelled protein for binding to a specific binding agent, for example, an antibody specifically reactive with a particular deletion polypeptide. The binding agent is, e.g., bound to a solid surface to produce separation of bound labelled polypeptide from the unbound labelled polypeptide. Alternately, the competitive binding assay may be conducted in liquid phase and any of a variety of techniques known in the art may be used to separate the bound labelled protein from the unbound labelled protein. Following separation, the amount of bound labeled protein is determined. The amount of polypeptide present in the sample is inversely proportional to the amount of labelled polypeptide binding.

5

10

15

20

25

30

Alternatively, a homogenous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding agent. This alteration in the labelled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the polypeptide.

Deletion polypeptides may also be detected by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which is also an antibody, and which binds the protein at a different site, is labelled. After binding at both sites on the protein, the unbound labelled binding agent is removed and the labelled binding agent bound to the solid phase is measured. The amount of labelled binding agent bound is directly proportional to the amount of polypeptide in the sample.

Western blot analysis can be used to determine the presence of a deletion polypeptide in a sample. Electrophoresis is carried out, for example, on a bacterial sample suspected of containing the deletion polypeptide. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support such as a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody is labelled, or alternatively may be it is detected by subsequent incubation with a second labelled antibody that binds the primary antibody.

10

15

20

25

30

The immunoassay formats described above employ labelled assay components. The label can be in a variety of forms as described above. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with polypeptides include competitive and noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant deletion polypeptide as described above. Other sources of polypeptides, including isolated or partially purified naturally occurring protein, can also be used. Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labelled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can be also be used similar to those described above for the measurement of deletion polypeptides.

II. Preparation of Deletion-Containing Mycobacteria

Mycobacteria containing specific deletions may be prepared by using methods of homologous recombination well known to those of skill in the art. In brief, homologous recombination is a natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar (i.e. "homologous") sequences, and the ligation of the two molecules such that one region of each initially

10

15

20

25

30

present molecule is now ligated to a region of the other initially present molecule (Sedivy, *Bio/Technol.*, 6: 1192-1196 (1988).

Homologous recombination is exploited by a number of various methods of "gene targeting" well known to those of skill in the art. (see, for example, Mansour et al. Nature, 336: 348-352 (1988); Capecchi Trends Genet. 5: 70-76 (1989); Capecchi Science 244: 1288-1292 (1989); Capecchi et al. pages 45-52 In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman et al. Cell 56: 145-147 (1989)). Some approaches focus on increasing the frequency of recombination between two DNA molecules by treating the introduced DNA with agents which stimulate recombination (e.g. trimethylpsoralen, UV light, etc.), however, most approaches utilize various combinations of selectable markers to facilitate isolation of the transformed cells.

One such selection method is termed positive/negative selection (PNS) (Thomas and Cappechi Cell 51: 503-512 (1987)). This method involves the use of two selectable markers: one a positive selection marker such as the bacterial gene for neomycin resistance (neo'); the other a negative selection marker such as the herpes virus thymidine kinase (tk) gene. Neo' confers resistance to the drug G-418, while herpes tk renders cells sensitive to the nucleoside analog gangcyclovir (GANC) or 1-(2-deoxy-2-fluoro-b-d-arabinofuranosyl)-5-iodouracil (FIAU). The DNA encoding the positive selection marker in the transgene (e.g. neo'') is generally linked to an expression regulation sequence that allows for its independent transcription in mycobacteria. It is flanked by first and second sequence portions of at least a part of the deletion or deletion flanking sequences.

These first and second sequence portions target the transgene to a specific nucleotide sequence. A second independent expression unit capable of producing the expression product for a negative selection marker, e.g. for herpes virus tk is positioned adjacent to or in close proximity to the distal end of the first or second portions of the first DNA sequence. Upon transfection, some of the mycobacteria incorporate the transgene by random integration, others by homologous recombination between the endogenous allele and sequences in the transgene. As a result, one copy of the targeted nucleic acid is disrupted by homologous recombination with the-transgene with simultaneous loss of the sequence encoding herpes tk gene. Random integrants, which

10

15

20

25

30

occur via the ends of the transgene, contain herpes tk and remain sensitive to GANC or FIAU. Therefore, selection, either sequentially or simultaneously with G418 and GANC enriches for transfected mycobacteria containing the transgene integrated into the genome by homologous recombination.

Methods of homologous recombination in mycobacteria are described in greater detail by Ganjam et al. Proc. Natl. Acad. Sci. USA, 88: 5433-5437 (1991) and Aldovini et al., J. Bacteriol., 175: 7282-7289 (1993) which are incorporated herein by reference.

III. Screening for Drug Susceptibility/Therapeutics

The expression products of the open reading frames in the BCGa1, BCGa2, and BCGa3 deletions of the present invention are targets for anti-mycobacterial drugs. To determine particularly suitable drug targets, open reading frames and surrounding expression control sequences are introduced into avirulent strains of mycobacteria, alone or in combination with other open reading frame regions to determine which regions are critical for virulence. Once particular genes are identified as critical for virulence, anti-mycobacterial agents are designed to inhibit expression of the critical genes, or to attack the critical gene products. For instance, antibodies are generated against the critical gene products and used as prophylactic or therapeutic agents. Alternatively, small molecules can be screened for the ability to selectively inhibit expression of the critical gene products, e.g., using recombinant expression systems which include the gene's endogenous promoter. These small molecules are then used as therapeutics, or prophylactic agents to inhibit mycobacterial virulence.

In another embodiment, anti-mycobacterial agents which render a virulent mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium.

Such transformed mycobacteria are useful as vaccine components, and as components of immunological infectivity assays. For instance, an animal's blood can be monitored for the presence of anti-mycobacterial antibodies using the procedures described herein, using transformed avirulent mycobacterial components in various

WO 96/25519

5

10

15

20

25

30

immunological assays. Anti-mycobacterial agents useful in this invention include, without limitation, antisense genes, ribozymes, decoy genes, transdominant proteins and suicide genes.

An antisense nucleic acid is a nucleic acid that, upon expression, hybridizes to a particular mRNA molecule, to a transcriptional promoter or to the sense strand of a gene. By hybridizing, the antisense nucleic acid interferes with the transcription of a complementary DNA, the translation of an mRNA, or the function of a catalytic RNA. Antisense molecules useful in this invention include those that hybridize to gene transcripts in the region of the deletions of the invention, particularly deletion region 1.

A ribozyme is a catalytic RNA molecule that cleaves other RNA molecules having particular nucleic acid sequences. Ribozymes useful in this invention are those that cleave deletion gene transcripts. Examples include hairpin and hammerhead ribozymes.

A decoy nucleic acid is a nucleic acid having a sequence recognized by a regulatory DNA binding protein (i.e., a transcription factor). Upon expression, the transcription factor binds to the decoy nucleic acid, rather than to its natural target in the genome. Useful decoy nucleic acid sequences include any sequence to which a transcription factor binds in the deletion regions of the present invention.

A transdominant protein is a protein whose phenotype, when supplied by transcomplementation, will overcome the effect of the native form of the protein. For instance, an avirulent mycobacterium can be rendered virulent by introducing transdominant proteins from deletion region 1.

A suicide gene produces a product which is cytotoxic. In the vectors of the present invention, a suicide gene is operably linked to an inducible expression control sequences which is stimulated upon infection of a cell by a mycobacterium.

IV. Use of Expressed "Deletion Proteins" in a Vaccine

The deletion polypeptides encoded by the open reading frames in BCGa1, BCGa2, and BCGa3 may be recombinantly expressed and used as components of immunological assays as described above or in vaccines. Expression of polypeptides

10

15

20

25

30

encoded by the open reading frames of the BCGa1, BCGa2, or BCGa3 deletions may be accomplished by means well known to those of skill in the art.

In brief, the expression of natural or synthetic nucleic acids encoding deletion polypeptides will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of polynucleotide sequence encoding deletion polypeptides.

To obtain high level expression of a cloned gene, such as those polynucleotide sequences encoding deletion polypeptides, it is desirable to construct expression plasmids which contain, at the minimum, a promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, i.e., shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. For detailed techniques employed in the recombinant expression of deletion proteins see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual (2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory (1989)), Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques (Berger and Kimmel (eds.), San Diego: Academic Press, Inc. (1987)), or Current Protocols in Molecular Biology, (Ausubel, et al. (eds.), Greene Publishing and Wiley-Interscience, New York (1987), all of which are incorporated herein by reference.

The expressed deletion polypeptides may be used in a variety of assays. For example, the deletion polypeptides can be used as reagents in immunoblot assays to test whether a patient was previously exposed to virulent mycobacteria (i.e., to test whether the patient has antibodies to the deletion polypeptide). These assays have the advantage of discriminating between previous exposure to an avirulent mycobacterium (e.g., one used in a vaccine) and exposure to a virulent mycobacterium. Thus, vaccinated individuals can be tested for antibodies to the virulent mycobacterium without regard to whether the patient has been vaccinated with an avirulent mycobacterium.

10

15

20

25

30

The deletion polypeptides can also be used as antigenic vaccine components to direct antibodies to elements which are critical for virulence. These polypeptides can be added to existing vaccines (e.g., those based upon avirulent mycobacteria and which lack the deletion polypeptide) to supplement the range of antigenicity conferred by the vaccine, or they may be used apart from other mycobacterial antigens. The vaccines of the invention contain as an active ingredient an immunogenically effective amount of a deletion polypeptide or of a recombinant vector which includes the deletion polypeptide. The immune response can include the generation of antibodies; activation of cytotoxic T lymphocytes (CTL) against cells presenting peptides derived from the polypeptides or other mechanisms well known in the art. See e.g. Paul Fundamental Immunology Third Edition published by Raven press New York (incorporated herein by reference) for a description of immune response. Useful carriers are well known in the art, and include, for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, and polyamino acids such as poly(D-lysine:D-glutamic acid). The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The compositions are suitable for single administrations or a series of administrations. When given as a series, inoculations subsequent to the initial administration are given to boost the immune response and are typically referred to as booster inoculations.

The vaccine compositions of the invention are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration that comprise a solution of the agents described above dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile

~ z'

5

10

15

20

25

30

solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient and more preferably at a concentration of 25%-75%.

For aerosol administration, the polypeptides are preferably supplied in finely divided form along with a surfactant and propellant. The surfactant should be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

The amount of vaccine administered to the patient will vary depending upon the composition being administered, the physiological state of the patient and the manner of administration.

Live attenuated recombinant viruses which include the deletion polypeptide, such as recombinant vaccinia or adenovirus vectors, are convenient alternatives as vaccines because they are inexpensive to produce and are easily transported and administered. Vaccinia vectors and methods useful in immunization protocols are described, for example, in U.S. Patent No. 4,722,848, incorporated herein by reference.

Deletion sequences and subsequences of this invention may also be used in methods of genetic immunization. Briefly, genetic immunization involves transfecting

10

15

20

25

30

cells in vivo with nucleic acids encoding pathogen specific antigens. The transformed host cells then express the antigen thereby stimulating the host immune system.

In the present invention, antigen-encoding deletion region sequences are used to transform mammalian host cells thereby resulting in the expression of the antigen by the host. This provokes an immune response by the host against the expressed antigen thereby conferring immunity on the host. Methods of genetic immunization are well known to those of skill in the art (see, e.g., Wang et al. Proc. Natl. Acad. Sci. USA, 90: 4156-4160 (1993); Ulmer et al., Science, 259: 1745-1749 (1993); Fynan et al. DNA Cell Biol., 12: 785-789 (1993); Fynan et al. Proc. Natl. Acad. Sci. USA, 90: 11478-11482 (1993); Robinson et al. Vaccine, 11: 957-960 (1993); and Martinon et al. Eur. J. Immunol., 23: 1719-1722 (1993), which are incorporated herein by reference.

VI. Use of Promoters within Deletion Sequences for Expression of Recombinant Proteins

Bacille Calmette-Guérin (BCG) contains all three deletions (BCGa1, BCGa2, and BCGa3) and yet is able to grow and reproduce indicating that the sequences contained within the deletion are not essential for bacterial viability. These deletion regions therefore make good target sites for the insertion of heterologous DNA as mycobacteria are tolerant of disruption of the native genome in these regions. The BCGa1, BCGa2, and BCGa3 deletion regions therefore provide suitable target sites for the incorporation of expression cassettes and the subsequent expression of exogenous gene products. The expression cassettes typically comprise a nucleic acid sequence under the control of a promoter. The promoter may be either constitutive or inducible. The cassette may additionally comprise a selectable marker such as an antibiotic resistance gene, a gene encoding a fluorescent marker (e.g. green fluorescent protein), or a gene encoding an enzymatic marker (e.g. B-galactosidase).

Alternatively, genes under the control of endogenous promoters may be used as well. In one embodiment, reporter genes under the control of endogenous promoters found within the deletion sequences may be inserted at the deletion sites. These reporter genes may be utilized as an assay for antimycobacterial compounds that act by inhibiting transcription or translation of deletion sequences. Assaying for the

.

H?

5

10

15

20

25

30

reporter gene product in the presence of an antimycobacterial compound provides a measure of efficacy of that compound in upregulating or downregulating deletion sequence genes. Methods of use of mycobacterial reporter gene assays to screen for drug activity are described by Cooksey et al., Antimicrob. Agents Chemother., 37: 1348-1352 (1993), and Jacobs et al., Science, 260: 819-822 (1993) which are incorporated herein by reference.

EXAMPLES

The following examples are offered by way of illustration, not by way of limitation.

Example 1

Identification of Virulence-Attenuating Deletions

Bacterial Culture

All strains of Mycobacteria used in this study were maintained in 7H9 (Difco, Detroit Michigan, USA) media supplemented with OADC (BBL) or were grown on 7H11 agar supplemented with oleic acid albumin dextrose complex (OADC). Escherichia coli (strain DH5 α or NM554) was used as a host for all recombinant plasmids and cosmids. E. coli was maintained in LB medium with or without agar. Carbenicillin (100 μ g/ml) was used in place of ampicillin for the selection of all E. coli plasmids.

Extraction of High Molecular Weight DNA

High molecular weight chromosomal DNA was prepared by diluting a late log phase culture of the respective mycobacterium 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing growth for 4 to 5 days. The cells were then harvested by centrifugation, washed once in TE (pH 8.0) and resuspended in 4 ml of 25% sucrose in 10X TE. 100 μ g of lysozyme was added and the preparation was incubated at 37°C for 2 hr followed by the addition of 100 μ g of proteinase K and sarkosyl to a concentration of 1% weight/volume. Following overnight incubation at 65°C the mixture was extracted 4 times with chloroform isoamyl alcohol 24:1, once with phenol/chloroform (1:1), and twice again with chloroform isoamyl alcohol. The resulting high molecular weight DNA was then run on a CsCl gradient as described by

Hull et al. Infect. Immun., 33: 933-938 (1981), which is incorporated herein by reference, and subsequently dialyzed against 4 changes of TE. BCG DNA was physically sheared by passage through a 22 gauge needle until an average size of 3-10 kb was obtained (20-25 passages). This DNA was then biotinylated using photobiotin (Clonetech, Palo Alto, California, USA) according to the method of Straus and Ausubel, Proc. Natl. Acad. Sci. USA, 87: 1889-1893 (1990), which is incorporated herein by reference.

DNA Subtraction

5

10

15

20

25

30

DNA subtraction was carried out between virulent M. tuberculosis H37Rv and avirulent BCG. H37R chromosomal DNA was selected because it was the most readily available chromosomal DNA from a virulent strain. In addition, M. bovis and M ruberculosis H37Rv are highly homologous.

M. bovis/M. tuberculosis specific probes were generated by the method of Straus and Ausubel, supra. with the following modifications. Sheared and biotinylated BCG DNA was used in a 10:1 excess for each round of subtraction. Wild type M. tuberculosis H37Rv DNA was digested with Sau3A to an average size of 1 kb. Hybridization conditions were 1M NaCl and 65 °C for 18 hours. Following five cycles (successive denaturation and reassociations) of subtraction, Sau3A1 adaptors (GACACTCTCGAGACATCACCGTCC and GATCGGACGGTGATGTCTCGAGAGTG were ligated to the subtraction product and amplified in a PCR reaction for 35 cycles (30 sec at 95°C, 30 sec at 55°C, and 3 min at 72°C). The M. suberculosis/M. bovis specific probes were radiolabeled by using one strand of the adaptor (GACACTCTCGAGACATCACCGTCC) as a primer and labeling with ³²P dCTP using the Klenow fragment of DNA polymerase.

An M. bovis cosmid library was constructed in the BamH1 site of sCOS (Stratagene, La Jolla California, USA) with subsequent in vitro packaging and infection of E. coli strain NM554 (Stratagene). 600 colonies were picked to Nytran circular membranes and the membranes prepared according to the method of Grunstein and Hogness, Proc. Natl. Acad. Sci. USA, 72: 3961 (1975), which is incorporated herein by reference. These filters were then probed using the BCG subtracted probe and positive clones selected for further analysis. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly, Nucleic Acids. Res., 7: 1513 (1973) which is

10

15

20

25

30

incorporated herein by reference. Restriction fragments that hybridize with the MTB/MBV specific probe were further subcloned into pGEM7z or pGEM5z (Promega, Madison, Wisconsin, USA) for deletion analysis.

Plasmid DNA for DNA sequencing was prepared using Qiagen minicolumns (Qiagen Inc. Chatsworth California, USA) and sequenced by the method of Henikoff, Gene, 28: 351-359 (1984), which is incorporated herein by reference, using the Erase A Base System (Promega). DNA sequencing reactions were run using a Perkin Elmer 9600 thermocycler and analyzed on an automated ABI sequencer. Analysis and assembly of contiguous DNA sequence was done using the ABI analysis software and SeQuencher sequence analysis software by Gene Clones Corp (Ann Arbor, Michigan, USA).

Deletion Region 1 (BCGA1)

Sequence analysis of over 16 kb of MBV region 1 and homologous regions in BCG revealed the precise junctions for the deletion in BCG. Eight open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 4). The potential start and stop codons and predicted maximum protein coding capacity are listed in Figure 4. Consensus ribosomal binding site sequences were found near potential start codons for seven of eight open reading frames. TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for 3 of 8 open reading frames in region 1.

Most notable is the ORF1C homology to an unpublished and uncharacterized sequence listed in Genbank as M. tuberculosis antigen esat6. A 65 base pair repeated overlapping (repeated ~2 1/2 times) sequence was also recognized within the ORF1C (esat6) open reading frame. Also noteworthy are the significant homologies identified between ORF1H and bacterial serine proteases including B. subtilus subtilisin. Of the eight recognized open reading frames, four (ORFs 1B, 1C, 1D, and 1E) are located entirely within the 9 kb region deleted in BCG. One ORF traverses the BCG deletion junction in virulent M. bovis.

DNA probes from the 9 kb deletion in region 1 demonstrated that this region is absent in all BCG substrains and present in all virulent MBV and MTB strains tested. Furthermore, restriction fragment patterns observed in Southern blot analysis

10

15

20

25

30

with region 1 probes are non-polymorphic and identical in virulent MBV and MTB. This region has far fewer direct and indirect repeats than the regions 2 (BCG Δ 2) and 3 (BCG Δ 3) characterized below.

The sequence of a small region, estimated to be less than 20 bp between basepair coordinates 10654 and 10664 in region 1 has been recalcitrant to automated sequencing. Therefore, pending sequence confirmation, the base pair coordinates given in the region 1 map (Figure 4) are approximations. The precise sequence determination is likely to effect the Orf1E open reading frame.

Deletion Region 2 (BCGA2)

Sequence analysis of over 15 kb of MBV region 2 and homologous regions in BCG revealed the precise junctions for an 11 kb deletion in BCG. Thirteen open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 5). The potential start and stop codons and predicted maximum protein coding capacity are also shown in Figure 5. Candidate consensus sequences resembling ribosomal binding sites were found near potential start codons for eight open reading frames. Of the thirteen open reading frames recognized in BCGA2, nine are located entirely within the 11 kb region deleted in most BCG strains while ORF2B2 and ORF2I traverse the deletion junctions.

TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for five open reading frames in BCGa2. A protein encoded by ORF2C exhibits striking similarity to the E. coli iciA protein which is thought to play a role in inhibiting and regulating the initiation of chromosomal replication. The iciA protein product is a member of the large LysR family of transcriptional regulatory proteins. Orf2F is highly homologous to an S. typhimurium ribonucleotide diphosphate reductase and a region of the E. coli and S. typhimurium proUVWX operon. Orf2H was found to have significant homology to E. coli and S. typhimurium permeases involved in aromatic amino acid transport and a eukaryotic cell retroviral receptor.

The Orf2G encoded protein was identical to the MTB mpt64 gene previously thought to encode a secreted antigen which is specifically expressed by MTB

10

15

20

25

30

and not BCG strains. Recent analysis of mpt64 expression revealed that three BCG substrains do express mpt64 (Moreau, Tokyo, Russian). Probes specific for mpt64 or other non-repetitive parts of region 2 hybridized to all MTB strains tested and the same three BCG substrains shown to express mpt64. Of interest is the finding that these three BCG substrains are derived from the original Pasteur strain prior to 1925. The current Pasteur strain and all strains derived from the original Pasteur strain after 1925, including the Connaught strain used in the subtractive analysis in this study, are deleted in the 11 kb DNA segment contained within BCGA2. These data indicate that an additional mutational event deleting the 11 kb segment of region 2, occurred in the BCG Pasteur strain sometime after 1925.

Southern blot analysis with probes from different segments of region 2 revealed a repetitive element located within a 2 kb segment (8-10 kb) of region 2. This repetitive element is ubiquitous in all tubercle bacilli tested. This element provides a marker suitable for RFLP analysis of mycobacterial strains.

Deletion Region 3 (BCGA3)

Sequence analysis of the almost 11 kb region 3 sequence and comparison to a homologous region in BCG precisely identified the deletion junctions for BCG. Twelve potential open reading frames were recognized in the region 3 sequence, seven of which are entirely located within the 9 kb region deleted in BCG. At least 9 ORFs in BCGA3 exhibit codon usage preferences comparable to that of the tubercle bacilli. Sequence homology analysis of presumptive protein sequences encoded by six open reading frames in region 3 revealed highly significant homology to listed sequences. Orfs3B, 3D, and 3E exhibit homology to phage sequences, suggesting a phage derivation for 4 or more kb of DNA in region 3. Homology to putative open reading frames in two M. leprae cosmids was also observed including homology to a putative bid gene encoding a protein involved in biotin synthesis. Also of interest was homology between ORF3A and an MTB sequence (mce) associated with cell invasion and intracellular survival.

Southern blot analysis with segments of region 3 deleted in BCG revealed that prototype lab strains of virulent MBV and MTB all carry deletion region 3 DNA. However, clinical isolates from PHRI are highly polymorphic or deleted in region 3.

This region contains many large direct and indirect repeats and, as mentioned above, at least 2 ORFs are homologous to phage sequences including homology to DNA invertases or recombinases. The repetitive nature of this region and the possible presence of a DNA recombinase could explain the polymorphisms observed in this region.

5

The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, was recalcitrant to automated sequencing and remains to be determined. Therefore, the base pair coordinates given in the region 3 map (Figure 6) 3' to the 9kb marker are approximations. The precise sequence determination of region is likely to effect the length of open reading frames 3H and 3L.

10

15

The foregoing subtractive analysis identified 3 regions in virulent M. bovis and M. nuberculosis prototype strains which are deleted in the avirulent BCG strain. The deletion located in region 2 may not have arisen in the original BCG Pasteur strain as this region is only deleted in strains derived from the original Pasteur strain after 1925. Region 3 is present in virulent MTB and MBV lab prototype strains (H37Rv, Erdman) and is highly polymorphic and at least partially deleted in the majority of MTB clinical isolates tested. Region 1 is apparently conserved and intact in all virulent MBV and MTB strains tested to date while all avirulent BCG strains tested to date are missing approximately 9kb from region 1.

20

Example 2 Screening and Identification of an Avirulent Mycobacterium

The ³² P labeled subtraction probe obtained in Example 1, was used to probe EcoRI and BamHI restricted chromosomal DNAs from BCG Connaught, *Mycobacterium bovis*, and various strains of *Mycobacterium tuberculosis* in a Southern blot. The hybridization was performed at 70°C in 6X SSC overnight.

25

The resulting Southern blot is illustrated in Figure 8. The probe showed no labeling of BCG reflecting the presence of all three deletions, while the other strains were labeled.

30

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

1 2

WHAT IS CLAIMED IS:

- 1. A marker for an avirulent mycobacterium, said marker comprising a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGala, BCGala, BCGala, BCGalab, BCG
- 2. The marker of claim 1, wherein said marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG.
- 3. The marker of claim 2, wherein said marker comprises a subsequence of a nucleic acid where said nucleic acid is selected from the group consisting of BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3.
- 4. The marker of claim 2, wherein said marker is selected from the
 group consisting of BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b,
 BCGΔ1ab, BCGΔ2ab, BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.
 - 5. The marker of claim 2, wherein said marker comprises a nucleic acid having at least 90 percent sequence identity with a sequence selected from the group consisting of BCGala, BCGalb, BCGala, BCGala, BCGala, BCGalab, BCGalab,
- 1 6. The marker of claim 2, wherein said marker comprises a radioactive nucleotide probe.

1 The marker of claim 2, wherein said subsequence is a sequence 7. selected from an open reading frame of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 A polypeptide encoded by a subsequence of a deletion sequence 8. selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 1 The polypeptide of claim 8, wherein the subsequence is selected 9. from an open reading frame (ORF) of a deletion, said deletion being selected from the 2 3 group consisting of BCGa1, BCGa2, BCGa3. 1 An antibody that binds specifically to the polypeptide of claim 8. 10. 1 A recombinant cell comprising a first nucleic acid that hybridizes 11. under stringent conditions with a second nucleic acid or a complement of said second 2 nucleic acid where said second nucleic acid or complement of said second nucleic acid is 3 selected from the group consisting of BCGala, BCGalb, BCGala, BCGalb, BCGala, 4 BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. 5 1 The recombinant cell of claim 11, wherein the cell is a 12. 2 Mycobacterium. 1 The cell of claim 11, wherein the cell expresses a polypeptide 13. encoded by an intact open reading frame from BCGa1, BCGa2, and BCGa3. 2 1 The cell of claim 11, wherein said cell is a mycobacterium having 14. one or more deletions in the genomic regions selected from the group consisting of 2 BCGa1, BCGa2, and BCGa3, wherein said deletions result in the attenuation of an 3 otherwise virulent strain of mycobacterium and wherein said deletions are present in up 4 5 to two of said regions.

- 15. The mycobacterium of claim 14, wherein said deletions comprise a 1 deletion selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 16. A method of distinguishing between an attenuated and a virulent 1 mycobacterium, said method comprising detecting the presence or absence of a first 2 nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a 3 4 complement of said second nucleic acid where said second nucleic acid or complement of 5 said second nucleic acid is selected from the group consisting of BCGala, BCGalb. BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1. 6 7 BCG_{\(\Delta\)2}, and BCG_{\(\Delta\)3}. 1 17. The method of claim 16, wherein said first nucleic acid specifically 2 hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic 3 acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from 4 Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from 5 6 BCG. 1 18. The method of claim 17, wherein said first sequence is amplified 2 prior to detection. 1 19. The method of claim 17, wherein said first sequence is amplified 2 by the polymerase chain reaction. 1 20. A method of claim 17, wherein said detecting comprises a Southern 2 blot.
- 1 21. A method of claim 17, wherein said detecting comprises detecting a polypeptide encoded by said first nucleic acid.

2	22. The method of claim 21, wherein the polypeptide is encoded by an
3	intact open reading frame of a nucleotide sequence selected from the group consisting of
4	BCG _{\(\Delta\)} , BCG _{\(\Delta\)} 2, and BCG _{\(\Delta\)} 3.
1	23. The method of claim 21, wherein the polypeptide is visualized by
2	antibody hybridization.
1	24. A method for determining whether an attenuated or a virulent
2	Mycobacterium is present in a sample comprising:
3	providing a first nucleic acid that hybridizes under stringent conditions
4	with a second nucleic acid or a complement of said second nucleic acid where said
5	second nucleic acid or complement of said second nucleic acid is selected from the group
6	consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab,
7	BCG\(Delta 2\)ab, BCG\(Delta 3\)ab, BCG\(Delta 1\), BCG\(Delta 2\), and BCG\(Delta 3\); and
8	hybridizing said first nucleic acid to the biological sample.
1	25. The method of claim 24, wherein said first nucleic acid specifically
2	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3	acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	26. A method of producing an attenuated Mycobacterium species, said
2	method comprising deleting from the genomic DNA of a virulent mycobacterium a first
3	nucleic acid that specifically hybridizes under stringent conditions with a second nucleic
4	acid or a complement of said second nucleic acid where said second nucleic acid or
5	complement of said second nucleic acid is selected from the group consisting of BCGa1,
5	BCGa2, and BCGa3.

```
1100
                                                                                                                                                                                                                                                                                                                                       1400
                                                                                                                                                                                                                                                                                                                                                                                          1600
                                                                                                                                                                                                                                                                                                                                                                                                                                            1800
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               2000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CGATCCCCCC CCCCATTTGT GCCCGAAACA GTCCCCAGCA GGTCCCGTTC 2100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           2300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     GACCOCCITIC ACATCGGAGG GCATCCAATT OCTGGCTTCC 3000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              COTCOCCCO ACCTATICGC ANATCGACGA COGCOCCCCC COCCITCTICG 3100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             3400
                                                                                                                                                                                                                                   TGAATCOCGT OCGTACAAGA TGTOCCTGCC GCCGTTGACC AATCCGGTCC 1000
                                                                                                                                                                                                                                                                                    CCCCCCACA 1200
                                                                                                                                                                                                                                                                                                               OCCAATCGGT 1300
                                                                                                                                                                                                                                                                                                                                                                TCTCCGCCGAG 1500
                                                                                                                                                                                                                                                                                                                                                                                                                 GTCAATGAAA CCCAGATCGA CCGGATTACC CGCGAGATCC CGGCGAATCG 1700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                    GOGGENECTIC CEGGAGESTA TECACETICEA CGAACTEGAC CEGAACECES 1900
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                TOTOCTICACIO COCCICIOSOS TOTACOCICAN CAGOCÓTICO CTAGACICAGO 2200
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                GATOCCOCCO ATCOCACCOC TOCCCCCTT ATTOCCOOCG OCGCCAGATA 2400
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            COCTCOCCC CTCCCACAAT 2500
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     CCCTOCCCAG GCATTTCTCG TCTCCCCAGA CCCCAAAGAG 2600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             COCTOTAGCA GGACCCGAGC 2700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      AACOCITICI GIACCCCATA CAAATACACG GACCCAAGAA 2800
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  GCTCCACGC 2900
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          OCCCOCTCC 3200
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 OCCURCAG PICACCOCOC OCCHONACTO TOTOGGAGAA 3300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  TOCOTOCOG AGATOCOCOC CAACCACATO ACCOAGOCOG TECTTANGOC 3500
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               CCGTATICTICS AACCAGGCAG CCCTGGCAAT GGAGGTCTAC 3600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       GATCCCOCC CCACCCAGAG CACCACGAAC CCGATCTTCG 3700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CCCTCGGCCA ACTGGGTGAG ATGAGCGGCC CGATGCAGCA 3800
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           GCCGACGAGG AAGCCGCCCA GATGGGCCTG 3900
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   OCOCOCCTOS TOCOCOCOS GTCOCTACCT OCCOCACCTG 4000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CONTOCCCC COACCOCITCG COCADGACCG 4200
                                                                                                                                                         GCCATACATG CCGCCGGCG CACCCTCGA AACCAATGGT GAAGCCCCIAG 700
                                                                                                                                                                                800
                                                                                                                                                                                                          900
                              TOSTCACCO GATOCCCCAG GTCTTGACCG CAGAACTCCA
                                                      GAGCCCTICTIC COCCOTOCCC GAATACGAGA ACTACCOCGA ACCCOGTOCC
                                                                           ACTIGITICAM AGICACCEGO ACTICATEGO OCTOTICADE COGAICTICE
                                                                                                        CACACCOCCO GTGTTCCCAT CCACAAACTG GAGCCAAACC TGACATATCG
                                                                                                                                  CCCCTTTCTC
                                                                                                                                                                                GOGGLACOG TICTOGRAGA GOCGCCGACA COSTGACOCG COCCGGCGAC
                                                                                                                                                                                                        GACGATOCAA ACCCCACCCA TGAGGAGGAG CGGCGCCCAT GACTGCTGAA
                                                                                                                                                                                                                                                                                                                                                                                       CACGTCATCA TCTCCACGCC ACGCTGGACA GAGCTGAAGT
                                                                                                                                                                                                                                                                                                                                                                                                                                            CCARCITICA COCCIOCAC ACCOCCCATA ACCTOCITGA OGCGATCACC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CGAACCCCCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CTACADACCO COTCAACACA GOCCAAGACO COTOCGATGO
   TCAGGIGAAT CTCCTGCTCA CCGACTTCAA AGGTGGTTCA ACCTTCCTGG
                                                                                                                                                                                                                                                             TCTCCCCCT
                                                                                                                                                                                                                                                                                                                                       CGATCGCGAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CTGACGACOG COCAGCTACG CTCCCCTTCG TOCTOGAGCG GATTTCACGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            OTGATOCCOS COCTOCTOS COGATOSTOS OCGACOCTO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             COSTANTONC ANGMOSTIC COGCCACCO GOCCOGANG ACTITICCANC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    COCOCAGGAG OCAGGTAATT TCGAGCGGAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                CACCICIONECOS CICACOCICOS CICACOCICOS
                                                                                                                                                                                                                                                            CTOCOGNICA TOGATGAACC GCCCCCCAT CTACAGGATG
                                                                                                                                                                                                                                                                                                             GCTGATCTAT CTCCAAAACC TTCCACACGT COGTOCOCTA
                                                                                                                                                                                                                                                                                                                                    COCCARCOG MACCACCTT CAAGGAACAC CGAGTGGGCT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             COCCAGACOG ACCTICACOCC GOCTICACTICC CACATGCACA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               OCCACAACOC TCTOCACGOC GTGACGGCCG
                                                                                                                              CACCOGNODE OCAGTACATE ACCAACAAGG AGAGCGGTGT
                                                                                                                                                                                                                                                                                  CCOGGAAGTC GACGCTACTG CAGACGATGG TGATGTCCGC
                                                                                                                                                                                                                                                                                                                                                              ACCECCACGT CTTTCTGATC ATCGACGGAT GCCCCGGTTT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        COCACCTAAA TACCGCACGG CTCATGGCCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ACCATOGACA AGITICOTOGG COCCOCATTIC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             GOTTANGATT ATTITICATTICC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            COTTOCAGOG CCAGTOGCGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           COCCARCCCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    COCCOCCICT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CCCCTACCCT
                              GAROCCCARCC
                                                                                                                                                                                                                                                                                                                                                                                         GITCOCCIC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ACCOCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CCCANGCGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      GCTATAACCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ACCCAAGTGA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          GCALTGCCAC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   TOCACOCTICA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CONCONCINCO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ATTATTTCAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                OCTACCCACA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         GCACCOCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   CACCOCCOCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            TOCCCCCTCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    COCTCCACCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    AAGACCGATG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CCCANOCOCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       GTCGATCCTT
GAATTOCTIGO GCACCOTGAT COTOTOGOTO GTOCCAATGA CTCATCCAGA
                                               TORROGRECIA TERATECTEC GACAGOECOS GATGAAAGTE GGCGCGGCCG
                            TOTICACCIA CATGGCCGAG
                                                                         GACCTACCCC COCTGCCAAC OCTTTTCGTC GTCGTCGACG AGTTCGCCGA
                                                                                              OCGITCOGGC GTCCCTGAGG GTCCATCTGC TOCTGGCTAC CCAGTCGCTG
                                                                                                                         MATCOCATTO COCACCACCA OCTICTICATGA ATCCAAGGCG GTAATCOCCA
                                                                                                                                                     AGCACCTTCT ACATCAGTGG
                                                                                                                                                                                                                                                                              OCOCCOGCG OCAACATOGG TATTGGGGGC GCACCTCAAA
                                                                                                                                                                                                                                                                                                                                CCGAGCCCGA CAAGGTCAAC CGGGTGGTCG CAGAGATGCA AGCCGTCATG
                                                                                                                                                                                                                                                                                                                                                        GTACCOBCAG CTOCOTGACG ATCCAAGTCA ACCOGITICOS TCCGATCCAT
                                                                                                                                                                                                                                                                                                                                                                                   TICCCCGACC TIGAGGGGCA GCTTCAAGAT CTGGCCGCCC AGGGCTGGG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       GICCGACTAC COCACTCOCT COGAGATTCC GATCGGCTTG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  TITCGOTGCGG CCANATCCGC CANGACGACC ATTGCCCACG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ATOCTICOCOS ACTACCOSCTO GOOCCTOSTO GACOCOSTOS COGACACCOA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               AGGAATTCCC ATCCAGTGAG TTCAAGGTCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CANCANGTOT TEOCHOCACE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                AGGCCATGGC CACGACGCCG
                                                                                                                                                                          OCCCOSTICA ACAGACCACT AGACAAGCCG COCCCATTICA CAGGITTCACC
                                                                                                                                                                                                     OCACCATGA GGAGGAGCGG CGCCAACGGC CCGCGCCCCCC
                                                                                                                                                                                                                                                      OCTEATEGEE CETGATEGGE GACAACEEET GEGATTTGCE
                                                                                                                                                                                                                                                                                                         CCAACCTICA CITICIANTICE ATCEACCTAG GIGGCGGCGG
                                                                                                                                                                                                                                                                                                                                                                                                            CCACTACCTC COCACCAAGA TCCAAGTTCCG GCTTCGTCAC
                                                                                                                                                                                                                                                                                                                                                                                                                                    OCAGITOTICA TOGANAAGCA CCATCTGATG ATCGGCCTGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 COCAGATICOC TTCCCAGCAC ACCGAACAGG CACCTCCGGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ACTOGOGOTIC AACCTIGAAGA AOCOGOTITOCO OCOGACCGAC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            accordant ecoccada
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        TGAGCCAGGC TTACAAGGCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    GACCATCOOG TITTGTTTCCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    OCCTOBACTO GAOCTOCCAG CGACAAGGCG CTTGCOCCTG CAACGCCGAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           GATCOCOTTO ACCOAGATOG
                                                                                                                                                                                                                             OCACOCTOCO COAGGTTGTO CTOCACCAOC TCOCCACTOC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CGATCCCTCC CGACATTGCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     GOGGECCAT CAGGICTCCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                OCCIOCOCIONO COGICCACIO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       TOGGAGGIG TGATCACCAT GCTGTGGCAC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              TROOCCAGIT GCCCCCCCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   TCGCCGCTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CACCTCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 GATCAGGCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      GOCAGAGATG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       GAGNAGCTCG ACCCGATCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          AACCCCCTT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CCANTCCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           NGAGGACGAC TOCTGAGCTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           ACTICIACTICS ACGGCAGGTT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CCACACCCTT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         GTTGTTCACC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CTCCCTCCTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ACCTUATION
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   OCCADOCTOC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    NOTICE ACCUT
                        GAATGGAAAA OCTTCCGCAC ACTGCCGCTG
                                                                                                                                                   COCTUGGCA TOGAAGACCC GCTCAAGTTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                AATCGAGTTC GGGCAATGCT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            CCCTOOCAGE TEMACACEGG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              GCCACATOCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CATCATTOTC ACCTOTCAGA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         CCCCCTACAT CGAGCCTCCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             GGAAAAATG TCACATGATC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   GTGACCGGGC TOGTTCCCGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              OCTCCACCGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               CCCCCCCATG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CCATACACCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         TTCGCTATCA ACACCATCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     CACCETITIC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     OCTGACCCAG CCGCTGCAGC AGGTGACGTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CTGATGTCTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 actrocricce acacceatos
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           GACGAGGACG ACTOGGACGA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CTCCGGCGAC CTGAAAACCC AGATCGACCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             GAACCATECE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CICANGGTAAA CACACAAAGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            GTCGACGATT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 GCCGAGAAGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CCCAACACA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CCAACCCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     CCCCCCTTAA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CCCCACCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            CCCAAGACCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               CTTGCGGCGG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               GICCOCTICTIC
                                                                                                                                                                                                     GATCCAAACC
                                                                                                                                                                                                                                                        COCTICAACGA
                                                                                                                                                                                                                                                                                                                                                                                                                                        TOCOGGINGS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CCTACTGATC
                                                                                                                                                                                                                               CCCGAAGTAC
                                                                                                                                                                                                                                                                                AGACGTTTCC
                                                                                                                                                                                                                                                                                                         CACTCACCCC
                                                                                                                                                                                                                                                                                                                                                                                                             COCCIONACE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CCCCCCCTCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       COCCACCAGA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     CCCCTCAAGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            COTOCITICAG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      TOCOCTICCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 GFTCCFTTCG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         GTCATCCAGG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  TCAGCCCGGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           GTAGGCAAAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     OCTOMOCITOS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            AATGCATCGG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CCTAATAGGC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               COCTCCAATG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ACCCCACCCC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CACCAACTTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CAGGCCCCAGA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            GAATGCCCTC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               CTCCCCACCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      GCTCGTTGAC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               acceptance
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ATTITIOCCCA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       TCAACAAGAC
                                                                      30
                        2 2
                                                                                                5 5
                                                                                                                                                                                                                                                                                                                                                                                                                                                         801
                                                                                                                                                9
                                                                                                                                                                                                                                                                              100
                                                                                                                                                                                                                                                                                                                                                                                   501
                                                                                                                                                                                                                                                                                                                                                                                                                                    100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    901
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            2701
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          3001
                                                                                                                                                                                                                                                                                                                                200
                                                                                                                                                                                                                                                                                                                                                                                                            601
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   201
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    2401
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           503
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    1601
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       2801
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             3201
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       501
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    3601
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               101
                                                                                                                                                                          5
                                                                                                                                                                                                   8
                                                                                                                                                                                                                           9
                                                                                                                                                                                                                                                    8
                                                                                                                                                                                                                                                                                                       2
                                                                                                                                                                                                                                                                                                                                                          ₹
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                2901
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                3401
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          1078
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     3601
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             500
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      <u>8</u>
```

9000 5800 5900 6000 6100 6200 900 6400 6500 0099 6700 6800 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8000 **010**0 8200 8300 8400 8500 8600 8700 8800 8900 9200 9400 5600 5700 9000 9100 9300 5200 5300 5400 ACCICATTICA CCACCCCCC AGRICACTICGA TGACGCCGGG GTGGTCGACG GGTCACTIGCT GACTICTGGTG **OCTICACGACG** GAAACCCCGA COTCAOGATG COCKDOCCAG GOCCCOGTCG AGCAACCCGG CCTCGCCGCA GCGCAACTCC GGTCGGCGTG CCGAGCGACG ANGOTGANGA AGGIGANGCC CCAGNANCCG NAGOCCACGA AGCCGCCCAA CHANGGTOOG OCTOOCHAAA CCACOCTGAC AGCAGCGTTG GGGTCGACGT GCCCGAAACC TYCCYGATYCG GGTACKXCGA CAATYGGGCG CGACCATYGC CAATOCOGIC AATCTOGAAG TOCTOCCOGC ACCOGAATAC CTCCACTIGC TTACCCCCCA OCTUMENTO TROCCOCCO CACAAGITICO COCCOCCOCT ACCOCCOTICO ACCITATION CONTRATAR COCTATACOCO GICATACOCOG COCCOCTOC TCACCTCCCC COTOTOCOGA OCCUPATOCCIA COCTOTICAC OCCUAACTIC ATCATCTICAT **OCCUPATCACC** OCCCAGAGGG CAATTTCGCG CACCACACA ACCACTCCCT GACCAAGCTC CCACCGCCTT AACACOGGAT CCCAAACCTA CONCOCCIC CACCCGAACC COGACCTIGCA CCCCACGTAC CACCOTOTOC CGTCCCCAAG CCCTTCCCCC CCACCCCCC CCTCCACCCT COCCCCCCT COCINCOCTT TOCICATCAA TCACATCATG CCOOGAGAAC CCAATGTCCC AGTTAAAGAC TOTTOCTOCT ATOGOGATOC GOCGOTOCITG GGAAACTICGG CCTAGCTITGT OCTOGCCGTIT COTOCOCCC OCTCCGCAAT ATTRCTOCGA CCGCCTTCAG CAACAAGACC CACCITCIATG GTCAGCCAAG CACCOSTCCCC ACCACCATCA ACCCUTICCC ACCACAACGT CAATCOCCA ACCCCGCCC GCTCCGTCCA ICCTOCGICG AGGITITACA ACCTCGICIT GGCTGATTGT OCCUPAGGAC ACCACATTG COCCCOGAAC CGAGATTTCA GACGATECTG ACCODEAGAE GGATGACEGA TITTOGTACTE ANTAITYCETC AGGCCGCGT CCAATACTCG COCCAACOCC GAGITCOCGT AGAATAGCGA CAGTOCTTCG TITTCCGCCGG CGCCCGCATC ACCOMPANDE TOCKCOCCO COCCOCCACC COCACCOCCA ACCCCCATG CCCATCGCCG COCCCCCACC CAACCACCC ACACCTCCGA TOCCCATCGC CONDCCCCC ACCARCCCGA ATCACCCCCC COCCOCCAG CTCCCCCCG GAACGGAGCC CTCCCCAGCG GICCIGOCAA GIGTCICAAT CGACGGCGCA CAACAGGCGT CACCATTITCG ACAGGCCTGG ACGTCGTTGA GCGCACCTGC GTOCTTICACG AGTCACCTICA GTTYCGACCGC CCAGAGCGGC CACCTGGCCG AGTGCCTACT GTTOCTCGAT CCCGTCGCGA CCCCGGAGGC TACCAGCGAA GACCOCAGCA AACTOCCCAA GCAACTTICTC ATCGGATACG TETTITETACA CAGCETOGIG GTCGCGGGIT TGATCACGAC ACCACACGO CCACCTTTCC ATACCTTCCG COCCCGACAA ACCAMOCICE COOCOCATEA CEGOSCEAGE AGECEGOCOG ACCEACCAAA AACATGACAG ACCAGCAGTG CAACGCGCTG CAGAACCTGG TETEMORGET TOTCATGGGG GCCGACTACG CCCACAGACA CCGAACGAAA OCCUPANT CONTRACTOR ACCACCTOGA COCTTCGACT TCACCGTCGC GINGCOCITCO TROTIGOTICOS GITCOATORCI CACOTECEDO OCTOCTOTOG ATCACCORDG TOTACGACAC GGAGATCCOG CAATGCCTGA ACCOTTCGCC GTCCATCCCA GATTCGGTGG TAGCAGCAAT CCAACATGAA COCOCCOCC TOCKSTOCKS TOCKSTOCKSTO CHOSCACCCC GROTHGFTGC CAGACGATICA GTICANACCOC CCAACAGGCC GCCGCTCCAC GAAGTGACAA CGACTCCCCC ACCCTCAACA TCAACCTOGG CCTGTCACCC GACGAGAGT COCCOOCTICA TOTACTICGGC GOCHETICAT TETCACGAAT GCGCCCAAGC CCCAOCTGAA TCACCCCTAT CAACCCACTG CAGCOCAACG COCCCTGACT CGAACACCAT CCCCCCACCG AGCAAGCGGA GCCGCATCTA AACCACCCAC MODICATOR COCCOCTICAG GTCGATCCGT CATCTCCACG ACCAMAGAA PECCHECIT CCCACCCCTA ICTTCCCATA TOTALCACCC MCGCCMCC **GCAMGATGC** 0000000000 CCCCCAGCC ATTACCOCCG ACACTAGCGT TTCCAACACA COCCATCOCC ACAGGTTCTA **OCCGACCIC** TCATTCCCAT CCATCCCCGT restendence TCCCCCACAC CCCATCCCT σσσσσσσο STOCTOOCCT TTGAGCGCAC MACTCCACCA CAAGGGGGCCG TTCCGAGGTG GICCCOGATE OCCOCCOCTG AACCTCGACC TIGITICACC ICTACCTICAC GGTAGCCCTG TOCTOGOAAA TOOGOTTICTIG ACCCGCTAAT CGTCCATTCA OCCTACCAGO GTGTCCAGCA AAAATGGGAC **STCACTOOGA** GTTTATACGT **GCGCAGCCGT** ACCTIGITCOCA **OCCOGNACCO** CCACCCCAAC CGACACCACA ACCCCCTCG TCCCGACGTG AGGAAGCATC GACAGAACCT ACCTICATTICA TTCACCCCAA CONCINCTAT CACATOCCCG TCGTCGGTCT AGCTISTICOCA CTACAACGAC ATCCGCGCAC TCATCCCCGA COGNETICON COCCATOCG TCGTGGTCAT CTACAAGCGC AAGGTCCTCG AATTGGCCGC AGCCCTATCC ATGTCATCGA CCITCATCOCS TTCCTCCCCA 10000000 TCATGAGTTG OCATACCAGO ACTOOCITYCE COCOGGGGGG ATCCCATTCG ACAACGAGGA CCCCCTCACC CCCCCCACT ATCCCTCCCA CAGAACATCG **AACACGTATA** CCCTCAAAA GCCCACCTC GGATCCAGGC CCACCCOOST CCTTOCTOOC CCTCCCCCCC CCCTCAAACC TOCATTICOCC STEACAGGTC ACCACCCAGC ANACCOLATO CGAGTCCCCT **CCTIGHTSCT** CCANGITICAL CCCGGCCGGG COCOCCOCC CTCCCAACCG OCTACCCOCC GANACCCTOC CTACCGACCG TTGGTCGACG TOCOCCOCCO ACOTOCCACO CCCCACCAGC CCAGGCGCGG CTGCCGGCAG OCCGACATTG OCTGCATGCG COCATCCTOS CTCTAGACOC GITACCAAGA TITICOCCAGC GACGACACCG TOCCGGTGCT GRACCATICAC COTICOTOGIC OCCIGATIOG COCITICCTICA ACCUNTAG CAGNIGCAGG COCUNTICCAG GGANATOTICA TOOCTTOGAC CGAAGCCAAC CCTTTCTCGT ARCTCCCCA CCATATCCCA OCCOCADENG AGCCGCCCTC **GCCCATC**GCC CCCAGACCAC CCCCCAACAC 200000000 **OCCUPACATE CODODOCO** AAATCCTTAA GGCCGGCGCC CACCACCC GACTCCATT CCACCOSTICTIC AACCGCTGCG CGGCCTGCCA OCOCCCOCC ATCCCCCTTT TGACCCCCCC TCGCTGTGCT GGTAGGCAGC COCAGOGGIG OCCUTACCOT COTCCOCCT TOTOCOTOCC AATTIGATICGA COCCOCCATG CTCCATCCCC TITTCCCCAG SOCIOCIC TOACCCAGTA CAGATGTCCC ATOCCCOCAC ACCTICCCAG CONCENTION COCCCAGCGG CONTROCTION COCARGEARCE CCAGCGCTGG CCCCCCCCC CCCCCACCC accracing TCGCTGCCCG **OCTICACTICAC** CONCETTICCE TOCCTICTICCE CACCCCCCAT GTTCCCCCCC CCCCCCCTCC GTAMATTTG cccccrccr **OCACACCCGT** CACCOCCTT ACCCCCCCCA ACCOGNOCC CCAACCAGTT TCCAMGAMGC CCACCCCTC GOGCOGTAG COSTTCGGAG **GCTCAGGCAA** GATGCCGATC GTCCATCCAT CCAGTCCGCC OCAATCCCCCG **OCCOCOCOCO** CCACANANAG CCCMCMCG ATCCTGGGGA TCCCAACCCC GACCCCCATC CACCTCCTCA **CCCCACCTCC** CCCCCCCAAG CAAGCTATOG AAACCACCCA CCGNATCCCA **CCCACATCAN** GANCCACCGA GTATCTOOCT CCATTTAGCC OCCACACAAG **OCTGACCCOC** ATTTCCAACA 2000000000 TOCCGATGGA AACTTATATT **accritecene** OCACCEMOCG CTTCATCACG COCCGATTCC CATCATCGCG CTCCCTCCCC TOCCHOOCHO CCAACTCTOG CCCANATTOC TCCAATCCCT CACCOCCCCA TOGATACCOS TOTAGACCOG OCCTOCCAGO ACCIOCCAGAC CAGCGCGGCT ACCOCCCCT CCCTTTACCC **COCCURCO** TCAGTCACTG ACCIDENTICE TATCTOCTCA Graciococt ACCACCACCA OCTATICGAGG CACCCAAGCC COCCCACTT CCGGCCCGCAC CCGAAGCCCA COCCAACTCC COCCCCACCC CCCACCCCAA CCTCGCACOG GICCCCCCCC GTACCAACTG **AACCGAGATC** COTCCGACCC CATCCTCGAC ACTOGRETICO TOOCTICAGGT TCATCTCCTT ACCTCCCCCC TCTTCCACCC CCACTOCTTG CTOCTOCOCC TOGACCCTAT **GCTCCTACCG** ACCCITCTOG TCACTCACTC ATCCCTTTICT **GCCGATTGGC** TOTTTTTCAC CTTCCCCTAT ATCCCCCTCC CCTGGCAGGC CACCCTGATT TITICOCTICOC ACCCCACTA **OCTCCTAAAA** ATCCGGTTCT TTCACCCTCT COCMOCOCC ATGCCAAGCA CGAAGACCCA TCAGGCTACC CAACTCGTTC CCCACCCCCC COCCACCAG **OCCEANDOOR** 5801 8101 8301 1096 5501 5701 6601 101 7301 7401 106 8201 8401 9001 500 8 5001 101 5201 \$401 5601 1069 6001 101 6201 6401 5501 5901 90 1201 7501 109 5 90 3001 3501 8601 8701 8801 106 9101 201 9301 80 5301 6301 5801 9401 501

12200 12400 12800 13000 13100 13200 13300 13400 13500 13600 13700 13800 13900 14000 14100 14200 14300 1400 14500 14600 12300 12500 12600 12700 12900 11600 11800 12000 12100 11400 11500 11700 AACCCCCCC 11900 10700 10800 10900 11000 11100 11200 11300 10500 10600 10300 10400 CCATCCGCTA GOCGGAACCA TCGCTGCGGC TGGCATCCCG GTACCGGGAG CCGATCCATG CICCITATAC GUGGACOCCG COOCCTOCGC GGACTTOCCG CCCCCTTGTC CCTGATGCGC GGCACCCATC GGCATTCCCA COGCUANTOT CACCAGCOCC ACCOGNITICOG COCCCCOAT ATAAGTCOGA TOTTCCCGCC TAGCCCACAC GTGCACCTGC GCGACATATT accrmoence ACCORCOGIG OCCACCOTOG CONTATOGOT TAGITTCGGC GAACTGTCCC COTATTCCCG CATOTTOTCG GCGGACAATA CCAGCTGTTG GGCGCGTTT IACCOCCATG COCCCGAGA TAGGTCCTCG CATTCCCCCA CGCCACCTTT CACCAATCCG GTCCCTAGTC CCTACCACCG CAACAATGCG TETTCACCIA ACCATETTEG ATCCCCAGGE COGTGACCIA CTCGTCCCCA **OCCURRENT COCHTICCCGC** TOCTICTIOCG GAGGGGGG GOCTCGATCT TGATGCCGCG GGGGAGGCTT COMPANDE CONTRACTOR TACTOCOCOT ACACOGGGG AATTTOGTOG CATCCCCTTC TECAGEOGO TEOCAGEETE CTUATCAACE TEGECATAGG CETTIGGEETE TCTTGCTGAT TATACCATT ATCGTCCCTA AACTGAAAGG TTCCTGCACT AATTTGATGC TCCCCACCTC ACCACGAACT CCAGTACCCC OCCCCTOCCT COCOCCCACC **acceccact** ACTOCOCCACA TCCCCCAATG ACCCCACACC ACTOCCCTCC CCCTCCTAAC TCTCCCTCCC ATACCOGGCC CCGCTCATCG **COTOCCOCCG** CAATATOCCT **OCCCOGTICAT** TTCCCAGTAA OCTORACTOC COCCORACIOS TOGACGACCT **GCACCCATC** ATTICITIOGAL TCAGCATCCA **OCCOCCOGAC** ATCCAACCCT ACCONCATOR CITICITICATE TGATTICAGES TECESTERS GOCCGAETGA ACCATACCOC TITICACCOCC ACCITOCCCO ACAGGOCTOC TENANGICIT CCACCCATCC GCANAGTUCG CGAGCGATOC COCCATCOCO GCCTCCACCT COTTOGCCCT GITCAAAATC AAGGCATGT AGACCGGGCA TCGGTTCACC GTCTCGCCGA TYCACOGTOG CCCCGACCAC AGGTYCTGAC TOCCOGTOTT GATCGAGCGA AGCCTCGCAA GCGGTAGCCG ATCHOGTAGC GOCTITICCTC GOGTOGOGAA ACCCOGCGAA CCCCCAAATA GCCGGCCGAA GTGGTCCCCG ACAAAGCCGA ACCCCACCAA ACCOCCACCA CCCCCCACCC TECAGGIECE GEOGITEGEE TCCCCCTCCA CCCTCAATOG GCCACCATCA ACCONTACC PACTOCICAL GOSCITICATE ACCTEGAACE ACACEATOTG GAGGTTTCCA TCATCCCCCT ATCCCCCTCA ACCITICCTCA ATCAACCCGG ATATTCCTOG OCATCCACCC COCOCCOUTA OCCUTIFICOC ACCOCOCOCOT PEGACTICETT ACTIGICETOG COCCGACGGT TACCAATGAC GOCCACGACG COCOCCAACC COCCGACACC GOGATCATCG ACACCGOCGT concensors CCCAACCCGG ACOTOMOTICO CASTOTOCOCO TICCTOCCAGO CCAATCTCCT GAACTGGCCA GCACCICCIONA AACCCCCTAG GTCCCCACAGC GTCCAATTTC CATTGGCGGC CTCATCCCCT CACCTGCCGC TOCTICATOGT THICTCCCC COCCINCOCOC TCACACCATA TOTOCOCTTG CCCOTCGACG ATGACCGTCG COCCUTAGNA AGTICACCOTIC OCCUMINACIÓ GACCICARGICA CCTANGGCCCG TAGGCCAGCC CATAGCTGTT TCGGACGCGT TGAGCGCCCGC CGCGATGCGT NAME TO CAST COUNTY OF CCCACCCCC CCCTTCCCAG ANCTOCT CCTGGTAATA ATUCCGCCCT ACCEGEATE TOCCCCCACT GTGACCATCG CCCCOCCCC ACCCOGACTC COCCCCTTGA TCATCGACCA TACCCACTCG ATCCCCAATG COCNOCACGA **OCTACCCOC** CCACCCCAA TOCOCTOOCC COCCOCTOC COCTICATIVE ACCAATGAAT ACCCAGGGCG TOCCCCTCGG CTCCCCCCCT GACCAGGACT STCACCTOCT COCCODOCOC TCCAGCACAA TCTTGGCCAC COCCULCCCC TATCOCATTT ACCCTITICCA **GIGTCOCCA** TOCATOOCOG CTTCCACCCC COCCICCOCC TTOCCGACGC ATCTCCTOGC CTCATCCCCC COCCCAGGC GTCGGGTATG COCHANGENIA CCTCCCCCTG TTATTOCCCA CTCCCATACC TTCACCCCCC CANTATTOCC COCATTOCOG CTGATCGGTA CCCCCCAGC COCACCACCT TCATCTTTOC TACATCOSTG **PCACCTCOCT** TCACATTGTG CCGCAGAATC ACCCGGTCAA OCCCATICAAC COOSCAAATT CCTCGAACTG COCCACCCC **CCCCCCCAT** CTAGGCATCG COCTOCOGTA ATCTCCTCC TOCCADCOCA CCOCAACOTG ACACCOCCOC CACCCAGCCG COTCTCCAMG ANTOCGATGA CATOCACCOC TTOCOCCATC AGCCOOCTOG ATCOCOCOGO ACATOOCCGA TAACGCAGCG CCCOCTGAGC COCCOCCTCG GCACCCAGCT CINCOCCITICO TEOCITOCOC GGAAGGATCC CCTCGGTGTA GAGCGCCTCG TCTTCCTGCT GAGAACCCTT CTCCCCTAAG GATCAATCCC TCGCTCAGCA TCGAGCCCGA CTATGTCTTC GAGTATCCAT GREGORITION AGCOTICAGGT AGTTCCCGTTC **GCCGTCGTTG** CACCITICCO CTCTTTGGCA CCGGCCGCCA TCCTTAGTGC GACATCGTCG AGATGCTGGC TCATCGAAGC TGCGGCCACA CACCGCGTCG ACACCATGGC COCADCECTA AAGGAITTET CATEGOCOGE TEAGCOCEEG TETAGCEAGE COCCITICATIO **acercacette** ACCICACICCC TOCCCCATG **GCCATCCACA** ATCAATTGAT ACCITCICCAA CACCATCTGT CCTCCCCCAT CTACCTCCCG **GCTGAATGAC** CATCCTCACT CCCCCACCGA CACCACCCCG recensoon COCTOCOTTIC CCGGTAGCGG CGAGCAAACG ATOCOGACCC COCCACACG CTOGACGCC **STCATCTOCC CCCCAGGCTT** ATTCCATCCC CCMATTICGIC COOCCUTICC **COCCOCCTICCA** COCCENETCA OCCODEAGOC COCCOCCATCA CICAROCCAGO TOCCATGIG AGATCCATGA CAGITICCTTC COCACATICAG TOCGTCATAT COGATCATCC COCCCTACTT CTGAGCCTCA CCCAGGTTGA GOCCGATGTC ATCOCCACCT GOCAGTOCCO GITGOCCACC ACCGTGGACT AGCCGCCAGT **OCCUATITIOS** CACCTOCTCA TACAGACTCA CCACGTCCGG CCACCAGTCC GOTTOTCCAGC GITCITICIAL GTAACCGAAC CTAAGACCAG GROCCITTIC TOCTODCOOC тесифелос соссеосите GTCCGGTTGG **GOCCTCGCCG OCCANCIAGE** TCTCCCGAAT STOGANACCC TTATTCOOCA COTAGICCOCC TOCAGTOCCC TCATCCCTCC CAGTOOCAGT CITCOCCEACE AGACTGACAT CCCCTTCACC CTOCAOCGTC ACTICACOCCO TGTCGGGCGC MCCACGCGT ACCAGTAGOG COCCTICATE OCOCOCTOE CCAGCGTCAA CGCCCGGGCG ACATGCGGGT **ACCUTCANTY** OCTICAGCAAC COGNICATIONS COCOCCOCCIT COCCEACCAG ANCTOCCT COCCOCCOCC ACCOCCCCA COCTOOCOCA COCTOOCAGC 2020202020 ACCICCACCCC **COCCUTATION** GTANAGCCGT ACCCCTTGAC COCTRICTING TCTCCCTCCA COCTICATO CCCAACCCAT CANGTACACC TITICICCATC STECCAGGIA ATCACCGIGG CGAGATAATC CAGTCCCCCA **OCHEANTENDE** ATTITIOCCOC CCCICATGIC CHOMOGRACC TOCOTOCOCO ATATCGATCT OCCUPATION 200200000 COCCTACCGA ACCIDENT GOCCAGOOCA ATCCCCGCCC AFFICTOCADE OCCITEGATICA OCCCATCOCA ACCCCCCGCA TOAGTTCGCA **GETICTICACGAC** TCCTTCGTTC OCCCANOCICC AACCOCCCC TCACCTGCGC ACCOGNICCA OCTITIOCOC CCCCCCCAT CACCTCCCCA TOCCOOCATC CCATGTGTGC TOTOCOCCGT TCANANCATT GACAATGCGT CTTCCACCGT OCCUPICAGAC **CCTTCCTTGC** CCNGTACAGE TGACATCCAC CCCCCCCCTC ACCOCCCCA TOCCATCOOD **GCTAGCCACC** GGACCITCATA ATTAACCAGC COCMACOCCC CCCCCTGCC COCTOCTICAC COCOGTGATC CCTCACCCGT ATTICTCCADO CTTCCAAGAC TICATOGGCT AGATICOCCTIC CCCCCCCAT ACCTCCATTA DODOCCOTICO ACCTGACCCG ACCATACATO GROCITOCITAG ACTCCTGCGA CTGTCCCCAT CCAACCGCC TTGATCGTCA CATCCCCGAT CCCCCATCCC CCAGCCGACG TCACACCCCA COCCOCITIT CCAGTTGTCA CCCCTTTCCA CTCCGACGGC CTTCCCCACC TTACCCCCC CCACCGTCAC COCCCGTTCA CCGATGCCGC ACCTICCTIGGG COCTCAATGT GICTIOGGCT GTOCACCGAC OCTACCGATC CCTTOTCCGA TOCCACCGCC COCCCCANG CACCTOCTT CCACCCCCCA TCCAGCACCC COCTACACTC CCCCTTGAAT CTTAGGCCAT AGAMGTGTC TCACCTATTG **NOCUTAGOCC** CCCCCTCCAT ATTICCGTGAT **OCAGACTICOC** CATAGGCCGC 10191 12501 12801 13201 13601 11901 12001 12101 12201 12301 12401 12601 12701 13901 13001 13101 13301 13401 13501 13701 13801 13901 1007 102 1301 350 4501 4601 4801 11601 11801 10601 1001 11201 1301 1401 11501 11701 10801 11101 10201 10301 10401 10501 10901 10701

16901 CSTANTOR	0001400000	ردور		
15001 Maranary	A CTATATURAL	COCCOCC	CCCACCAGGT CGTCGACCCC	acceceasas forefeces feecestear econtegate arctecees 15100
15101 (2727)	L GESTANORY	COMPOCE	COCCOCCAG	TATTOCCOCCO CANACCACOT: GCCGGCGATG GCTACCGCCC CCTCCCCGCC 15200
15201 PRICANCICA TENEGOSTE CA	A TTCACCOSTT	CCCCTGTC	OCCUMOCOC	CTOCODOCOC GOCCACOTICE ACCCACGOTIC COTOCCATICGA GAACGAGCTIG 15300
15301 COCATCCC	G TCTGGCCGAT	COCCCGACG	CTTAACACCA GCGGTGCGTA	CCACCACCACA STEAM TOTAL AND STATE STATE STATE STATE STATE STATE STATES S
15401 COCCIGIO	A COGNICCOC	COCATTOT	GTACOCAATC OCCACCOGTG	PROCESSECTO CENCEACEAC CACCACOCCY TITSACGITISA CCOCATAGIC 15500
15501 GATGGATGC	A CCCAGTGAGG	TCATCCAT	COOCCIOCIC ACCITICIAGE	ACCOCCUTIC ACTICATIONS ATCACACICCA COCCIGACION CICCICICION 15600
15601 ACCACGGC	C COCCANGACT	CCATCCAA	CCUSCOSCCG COGTGGCGTT	
15701 TCTGACGTA	D CCACACCAGT	MOCGICOG	OCCICCACGCC GACGAACCCG	reactivaces expacedace exceptuats satisficial gasteceats 15800
15801 COCATCACA	G TCAGACAGGC (TTACCOC	CTOSTCGACG AAATCGCCGC	CARCITICICA COCAACCCT CACCAACCT CACACCCCT CTCCATCACC 15900
15901 OCCACCOTIC	A cocoocce	TCCCGAAC	THOTOGGCAT COCCACOCC	CAGATACGTG TTOCTCCACG GCGGATCGTG GAACCCGGAC CCCGGCAGCG 16000
16001 TOCTOOCC	A CCCCCACAAA	CCCCTGTT	COGTAGOCTG ATCCGGGCCC	ACCANCETOGO ACCOCANACIO ACCOCAGATOS ATCOCOCATOS CICITAATICIC 16100
16101 CGATGCGG	C CACCOCCTCA (MCCCCAG	COCCACCOTO ATCAGAAGA	TACOGRECAC TECESGAMEA CTECATTEGT TEAGATTECAT TECGATTECAT 16200
16201 TCACCTCC	T TOCTACCTTG	CCACTTCA	COCACCTICTO TOCATTITAG	ACCTANCECE TEXTECANACA ACCEPTENCE OCCUPACATE STOCKCOCCE 16300
16301 CCGACCAGG	a cocotadoco (GTACCTOC	ACCACGCCGG GACTICAACGG	TITITICATACE CEACTACECE ATATECECENT CETACENANE GATECECXXC 16400
16401 ATGTCTCGG	I TOTCTGAGCA	COCTOCOT	ATCOCOCCAT CGATGTCGGT	OCCUPINATE ATCTICAGAT CCTGAACCGA TACCGGTTGG CCCGCACGTT 16500
16501 TITICCOCAA	C CACCCOGGIG ?	CCCCAACC	CTTCOCCCC TTCGATCACG	TICCOCCCC ACCONCION TICCATAGGG TCGATACCGT CCTCCCCACT 16600
16601 ACCIDENCE	G TACTTACCCA ?	GTOGTON	COCCITOCAGO AATACCITOCC	GIBCOBOCOTO ATOCAGOTOS CITOROGOCOS GIBTAGOSTA OCOGNOTOCA 16700
16701 ATCTCGACGA	TCTCCACCGG	CCANTANGAC	TOGANCOGCA OCTITICOGITI	GANCCOCICA CICCANACICCO COTTICACION GAIGGAATTICO GTACCCCIGIG 16800
16801 Trccaaarc		TCCGGAGAGC	TECENACIOG TTOCATOCAT	ADCITICACITÀ INCIDATACITO INCIDIO DE 16885
- 10	0 20	90	40 50	001 06 08 02 09

2900 3000 3100 3200 2200 2300 2400 2500 2600 2700 2800 3300 3400 3500 3600 3700 3800 1700 1900 2000 2100 CTCCTCCCGA ACCOTATGTG CCCATCGACG ATCACGTGGT CCAACCCCCG 4500 1300 1400 1600 1000 1100 1200 1800 900 800 TACATGACAA ATCAATAGAC AAAAGGAAGA CAGGCTGCCC ATGGGAGTAA ACCTICCIONA CTOCTICCIONA ACCIOCICA L'CAACIOCTIG TITROCOCTICA GACTUGICCA OCCITATUDAC OCACOCOSAA CAATTIGAGIG GICOGAAACG CAACCACTT CAACCACCC AGAGAAGATG CTCGAGGCGG CCTACCGACT AACCTCGGGA ACCCCTCGAGAACG ACGACATCTA TCCCGACGAC AGGIGINATION GINCOCOCO COCCCOCIUT COCCCICOTT GICACCOCOT CATCACTTAT AAGCCGATAA GCGACATTAT GTCAAGTGAA GCTUGTTA AGGCTTGCCT GGTGATGAAT GTACTTCGGC GAGACGCAGC ACCTCGCCAT CCTCGTACGG CCCAACGACG ATTITICIANCE COCCICCO COCCATCOCT CACCOCCCO GTCCAATACC CATAGAACGA AGTOTGCCAC TTCTAGCAAA GGTGGTACAC CTACTGGCGG GAGCACCOCG GAATCCACAT CCCCCCTCG CACGAACCCG CCCCTTCCA TTACATACGA GOSCTCOCTG CAGCACGAGA TCGGCCGCGA CAAGCCCGCT GOTTCOCCAC AACCCCCAAAC GGCCGGG GCAACCGGAT CGAGGCGGG TOCAACTECA COCEGACEAE CEGEGEEEE GETEGEACTA COGGACOGCA CTOCOCCOC OCTGACTACC COCOCTOCCC ATTCGTCATG COTACCOTCA CTOCOTAGAT GCCCACGCG CCGACCGTAG CCCGCCACCG COTCCGCATC AACTCCGAGA TGACGTTCGC CCACCTICAG TOCOCAGICG CGICTAGGIG CAAGGATATT GCCCGTIGAG TRAACCACCG GGGCAATGA TGCGATTCCA ATTCCCTGGG CANGACOGAT TETECTOCCO GCAAGTCGAA TTCAAGCTTC CAATCGCTTA ATOTOLOGIC CAGGACTAGE CTGGCCOTCG CCGCACTAGE PROCESSARIA COCTIGENCESC CACCOCCTICA COTAMATOTIT CAGECCOCCC ACCANCITING CITCAACCGCA ACCGIGATGA AATTITGGCCT GACCCCACTC CCCCGATTAG CCCCGTTCAC CGAATGTOCA CTOGTGOCGC CGACACOCCC GOGGGAGGCC GROCTEGGG GTCANDOCCG TTCGCATGCG GTCCGCCTAG CCGACGGCCA OCANGINCING GACCOCTITIG TGATCACCCG GCCCGAGCCG AAACTGAATG CCGTCCCCCC **OCCOCGINGAC** OCCUPANTIC OCAGOCAGET GGATICIOCO ACTOCIOTIT COGCOATOT TCCCCACK:TT ACAGAGGGAA CTIGTCGAGC CCTCCCCAGA GITICITICCCC GCGACCACAA COCGACTGCC ACGTGACTCG CTACTCGCCG GTCACGATCG ACCCCTUTTT **GCTCACGAGT** ATCCTCATCG CCAACCCCCT CCCCCCCGA CCCATCAAAC CACTITOCCCG GITTIGGACCA GCAATAGCGT CACTIGTGACG AAACAGCCGC GATCAGACGT TCCATCGGGT ATCGCGGCGT GCGAGTGAGC CACCAGCTOG ACCCCCAATT GCTCACOGGC CAGATCGACA OCCCTCTAGA GICCCCTTGC CACCGAGGAC GTGACGTTCC TOCTTGGTGT AAGCCCCCCC GCCGGGGGAA TACGAGGAAG TCGGTGACAT CGTCGATGCC TCGCCGCGTT TEGACCIOCE CTATOTECICE GACGECCTEG COGAGGICET AGAAACCATG GACGATGAGT GGTGATTAGT COCCCGGTGC TCGTAATCCA GTCAGATCCG TACAACGCAA GAGGIGCCAG CGAGCTTGAT GCACGAGGTT TETAGGCETG COTCGACTIOC GOCTIGGGGCA GCGTCGTGCT CTGACGCCTG CTCTTTACCA AGAACTTCGA GCCACGTAGC OCCIOCACCGA CCCTCTCCTC AAACCTTCCG GICAGCTAGA ACTICTOCCC CCTCACGTCC GOOCATECEST GREGTAGOOC TOCCCUTITIC CAGCCCCAGG GCTCTGGCGC **NGACAGCCTC** NOCANGCTAC COCCITCACGC TOCTOCOCOG COCCCCTOCC **POCCACOTICG OCCUPACOCC** ACCOCCIOCT COTOCAGIGT THETETROCOG CCCCAACCT CCGAGTTCCC ATCTTOCTGA TTCCGCCCAA GCCTGCATGG TTGAGGTGGA ACGGGCGGCG TOTOCOTOCT AGCAGCGTCC GCAGCAGGCG GGACGANATC CCGTACGGCG TECCENTICA CEACOGIAAT GECAGOGIAAT CONTEGEOGA GGACACEGAC CCCCACACOC TCTCCTCCC OGATICCTICOS ACTIGACCIOCO GICGIOCITIO TOCACGAGITI CACCGAGGIC GTTCOCCCC TOCCCCCCC TCCCCCTCA AGGCGCTCGG CTGCTGGCTC CCAGTAGACA TCOCAGTOGT CCCCCCCTCC TCCCCCCACA TCOCCTCCGA TCGAGCGCTT GGGATTGACC OCTICAGGTOG TTCCCCAACG CCTCAACAAG CAACGACTTA TTCCCCGAAT COSTTACTOS AACACCGCAA OCCOGRACION GIGITICOTOT GACCGATICCE CTCCATIGAAG GCACCOTICAC COCTITOTOCO GTGAGGGCGC ACCTOCCOT CCTCGATCTG ACAACCCCGC COGACATOCO CCCTOCITICT TOCACATOCA ATGAACAGCG AACTCGGCCG GOCCAACTAC AACTGCGACA AACCGTCGAT AGGATOCCOC TGTGCCATOG TCATCGTAAG GAGCAACTCG AATCAGGIAG TITTAGCCCG CTGCTAGCGC Gregaceate gatgateage egeegeaata TENCETACET CCGGGGTTTC GGTTATTTGG TAGCGCGGAC GAGTGGTCGA CCCCCACCAC CACOGTICAGO GTCAACCTTC OCOCOTTGCG AGAAGCGAAT TCCCCCCCC CGATGTCAAA CACGAGCTCA GCGGACACGA GCACTOCCAT CAATCCAATT TCTCOOTTCA GCCAACCTTC TGGTCATCAC ACCTOOCAGE AGCCOCATTG TCGCAGTCGG TCCATGTGTG GAMCCOCTC TGAACCAGAT ATCOGCCCOG CCGAACGCCA ACTTGCGGAC TCCGTAGGGC OCCUPICACIO TOTA PARA TORA STANDAR STOCAGACTO ACCIOCOCOCO CCCCCCCAAT COCITATACT ATCIGIATCA AGACAGCTAT **OCCCTACCTG** OCCUARGCAT TOTOLOGIC MICAGACTG ACCTCACCGA CCACGCGACA ATTCGTCACG CATTACCCC **CCCCTCGTTG** CTCGCATCCA COCCTCGGTT CGTGTCATAA GACGGCGGCG AGTAAGNACC GGTCAGCCCG CCTCTTAGGC CCOCCATCAG GCAGTCAGCC GGCGAAGCGC CCCCCAATCC TOCGCCGTTT GROCTAGCGG GCACGCTCCT CCOCCOCTIT ACCOCCCTG CCACTCACCG CACTTCCGGC CCCCTCCCCA CCCTCCCAOG AGTOCOTOCC AGATOSTOCA ACOGTOCTAT GTCATGCTCA ACCCCCCAT CCADGATOOC AACTATCAAT CAATGAGGG ATATCGCCC TOCARCACOA TTCTTCACCA AGCCTGCGCA **COCCTACCCA** CCTCCCCCTC TCACCTCCAA TACGGCGCTG CTTGAGGGCA AGGTTGCATT **COCCIOCOST** COTTCOCCIAC COSTATICAAC OCOCTACAGC **OCACTATOCC** TCGATGTTCA AACCGCAGGC TGACACTOCG CACCTCACCC TOCCACATOC GGCACGCAGC CTACCCGCGG AACGCGGATC CAGTOCCCCT CAGCCCCATG CACTAGACCC ATGCCCACTC ACACTANTOC COTCITICIOS ANGITATACC CAGCATCCAT GCCACCGACG GTGCCCCCCC CTCGACCGTG TOTOGATGTT ATACCOCCC **OCCAGOCOGT OCTCGAGITIT** OCCACCCANG ACACCTGATC TCCCTCCCAG TCACCTCCCC **OCNOCOSTICA** TCACCCCCAT TOTOCCCACC CTCCCCAAGC CTCCCCACGA TCCAGCGTGA TCAAGTTCCT ACTOGACCTT CACCCOCCTIC CCCATCGTCG GCCGCAGAGA **OCATIONOCCC** GITTOGGTGCC GACCCOCTOC CCTCACCACC CAGACAACTC TOCCATICAG TCTACTGGGC CACTIGITICATIC GTCCTCAACG Trespecen CACCTCCTCT ACCCCTTCCC GTCCACACTC CCCANGCCGG **COCCITICOTO** ATIGICACIC ATOCACCOCA CAGGCGAACG TCCACCCCAT TOCTACCOGT ACACACGATC TOGECAGGGT TTCCCTTACC GATGCCCCGT CTCACAACGT **OCCOCCCCOT** ATCACAGCCA 1002 2601 1062 3001 101 3201 3501 1076 108 1301 9 301 5 2501 101 3301 601 180 1901 001 1101 301 501 101 30 5 501 201 101 201 ₹ 201 80 555 8 9 9

6800 5300 5400 5500 5600 5700 5800 5900 0009 6100 6200 6300 6400 6500 0099 6700 6900 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8000 8100 8200 8300 8400 8500 5100 9600 8700 8800 8900 9000 9100 GACCACAGAG ACCTCGACCA OCCGACGACG ATCCTTOCTC 7007000 MCACCAGGA CCCCCCAACA GITTIGICITIC GAACACGAGC TCACCTOCOG GTGCGTCGAT ATCACACAGA ANCONCITIC COCANACCON CETTICANTITE COCOGCOTEG TEGOCTACTIC CCCCCCCTTCC **ACTIONNESS** CTCCCCTAAG CCCCCAAACT GAAGCCCACG **CCAGTGCGAC** CTTCCATGCG CAAAAGCCAT ACGCCTCCCT CGACGTTCGG COCOTOCACC **AAGCTCCGTC** cocarcinoco CAACCCCTCT TGCTGGACCC CAATTCAAGG CCCCCTCAAC TCCCCTTTT CTTGTGTGCT ACCCOGTATT CCACCATCGA GTCCACGCCG CCCTICITAACC COTOTOCIACO CCCCCTVSCC CONCINCOCC CCCGATCCCG TIGGOGATICAA CIGGOGGGAA CTCCTTCCCC **GCAACTIGTGA** GACCCCTFICA TACCTACATA CACACCAGCA TOCGTACCGA CCCTCATGTC CCACACTGTA OCCOCCCCAT CCCTCCCCCT ACCCCAATGA COCCTTCACC **ACACCCAGC** ACTITICICCC COCCAATTAC CONCACOGTG AGGCCGCCG CCCTCTTTCC TTCCCCTCTA CTOGAATCTG AAGOGCAAGA CCCACCOGAA TCGATGATCG AGGGACGGG TATGGGAGTC TCCCCAACCC CCACCTTCCC GAAGATGTAG TACAGCCCCC GCGGCACGCT TCTGCCGTCG CCTCGACGAT TCTCCACCAG TCACCAGGAC GAACCCCCC CACACCCGTGC AGCAGGCCAG TTATTAATGT AACTCGGCAG CCTCCTCACC ATCCCTCCCA ACCITCCTGCT **SCCCOCCTICC SCOOCOCTA** ATCCTTCCTA CICCCOCCT CITCCCACCG CTCCCTCCCC CTOCAGATGA CTCCACGCAG CCTTGTTCGC CCCCCCTT CACCTCCCC craccarra COCTCCCCTG TATACCTCTA AGACCATAAA GTATISTICAG TEGOGGATGA CAAGCCAGGC GCCGACATTC COTAGAGECG ATCACEGECG GOGCTGGTGT AGACCTCAAT COCCHOCIC CACCOCATIG CACCCAACTC CCAGAACCCC ACAACAAGAT CAGCGGATCC GACTECATEA CEGAGGEEGE OCCUDANTE CITCARCECEC CCCCCGGCCC GCCATTCGGA TOCOCOUNT CCTOCCCCCT CGAGGCCGCA CAGCACGCGT TCACCCCCTC CTCGACATCT **GCAGCGAAAT** TCTCACTTC CACACCCCCC CAGTCTCAGC **GCCCATAGCC** ACCOCACTOG CTGTCACCOC GCCGAGGCCG ACCCCATTAG COCCUATICAG CACGATIGTICG ATCAGCGTGA ATACCTAATA ATTICTTICATT TACATTAGGA **OCCUPATION GCTCGCTGAA COTOTOCOCC** CCAACCCCCT AGACCGACGC TTCCCCATOC COCCICCOC ACCIACATAA TCACCICACC **GCTACCGTTC TGACCTGGGG** COCATCCOTC TTCGGACATC CCCCCCTCCA **GCACCGTCGC** GIOCOTTOAG GAACOTCAAT OCCOCACAGG CACCAGGICG GCCAGGTGCT TOGICTCGAT TCCCCACACC AATTTCACTC **GCCAGATCAT** CACTGCCGCC CCCCCCCAT CACCCCCCC CAGATCGTCC MAGCGCTCG CCCCCTCCT GACCAGAAAT CCCCACGCCA TOCCOCOCO OCCACOTOR GOCACAACTT TOCOCCOC CCTOCCCCC COSTIGOCO CACCATOTIC GAACGACTIC ANGOGGGC ACCGCGCGC CCCAACTCCC CCCCCAACCC CATCATCACC GCGATCAGGC MICCOCCC COCCATICACC **OCATICITICO** CACCCCCCT AGTACCCGT CCCACGOCTT CCCATTOCCT AGTCCCAOCC COTOTTICGAC ACCACCGAGC **GCAAGTTOGC** TGTACACCOG OCAGGTAGCG GGTTGGGACC TAGGTCGAAA CTCAGGTGAG GATCCACACC CACATICCAGC COCCCOCCCT TATACCTOCC ATTATCTCCA CONTROCCCC TACACCACC OCTOCATTICA COCCOCCCAG CTCTTCATCC GCCCCCCCAA COCCMACCAC TCTCGGATTT TCAGTAGCTA TCACCTAGAC CCACTAGCCG COCCTACGAC **ICTOCOCCC** CCOCCCCCC DOCCOOOTC CCCATACCCC CACAGCTGGC GTCGTTGCAG **OCCUTIOCITO** CONTINUESCO COCCATCTAT CCTTTCGTCG GAAGCTOOCA CCGATCATCG GINGINGIE CCCACCTCCC ATCCCCTAAG ATCTCACCAT **CCCCCCATC** COCTTACCTC TOCCAGCCC PCCCCCATCC TCCCCACAT CTGTAGTCGA GOCCCTCAAC ACCONCOTOG CCCCCACATC TOTGACATGA CACCACTCAT DOCOCCCC **OCCUPACIO** TCTCCACCAA ATCAGCGCCA OCCACCOTOG CCCGTGCCAC GTCAAGAACT OCCATACGC **ANCANATICET** CCCCMCCAT GCAACAGGTC GCCACGCCCG OCCOMPCICE ACCAATCCCT AGCCCATCAA COCCOMMCA ATCCCCACAT GTCTCCCCAT ATTCGTCCAG **STOCKTANDS** TOTTCCCCCA ACCCCACCAT CCCTTCAACC **GPTCCGATGA** THICTCCACC ACCCCTCCCC GIGTTGAACG ACCTACACCT ATOCGTCAGT CITICACOCTC ACTOGACAGE CATCAATCCC AGAGTCCGCTC CCCCCCCAGT cocconcer TYCGTCGACT occoence CCCATGTTTC TTCCCGTTAC CCAATCCTCC GACGTTCCCC CCCCCACCAA CCCCAAGTCC CCCAGGATTC CCGTCGGCTC CTCGACCCC Teccocccc CCCCCCCCTC ccoccocca **AACCGTTOCC** COTTTCOOTT CAGTTCACCG CCGATACCTC TOCCOCOGAC **GCCTTTTGACA** TCAAAGACCC CITITIOCAG SCOCCOCAC GACCETTICT CHOOOCCACC MACCACTECT GGCCGTCGGC ACCCACATTC CTAGCAACAC GACGGTGTCG AATOCTCCAA coactecta Acceasage CCATCCCCAG **ACTISTICACTIC** CGATCAATCC OCCURRENCE CTGCCGCAGC COCCCATTICC CTACCCCACC CCACCACCC COGTICACCO THOOCETTOC CCACCACCAC **GCCTCAAA**CG TOCOGTICOC **GGACATICITIG CCATCCCCCA** CAGACCICAAA CCACCTCOCT **OCOCACCCOC** COGTGAGGCG COSTCATCCA GTICTOCOCC MATATTATOG TIGICCCTAT GTAATCGTTT GACCICCIATOC CCACCACCTC COCCATICGTC GETICTICGGAT CGATCCCATT GTCAAGCACT TTGGCATGTA ACCOCCOCTT CCCCCCTCCC GTCCCCGAAT **GCCCCAACAA** TCCTGACGC ANCHOCOCT OCCACCOCCG **OCCCTICANTA** TTCAACCCA CTTGACCGCC COCCCCCCCCC CCACATCCCC CCCAGCTGGC ACCOCTOCCC CATCACCAAG TOCAGACGCT **CCCACCCCAAC** OCTCACOGIC COCTTCACCC TOCCATCACC CCATGTGCCC ATCCACACTA CACCCCCTCG ATCCCCCTGT TTOCOGTAM CCCCCCCTG CTACCAGTOG ATCCCCTCCA COCCOOCIO CTCTATTGGC OCCOCCIONCO GACCOCATCO COCCCACCG **GCGCTGGTGA** CCCCCCCAOC CACCOTCCCA CCCAGCCCCC COSTCOCCC CATGACATAT CACCCOSTA OCCOCOCTCA **CCCCTTGACG** STOCTISTOCC ACCUITICCC CCCCCACCCC ATTAGCOCAC **OCTOGATTCC** CICACATOCT CTTAACGACT TTOCCCACGT TCTCCAGCAT TCAAAGACTG ATCCCCCATC GAACCOCCETT CCACCCCAT TACAACAGGC CCCCATACGA ATACAGCGTA TCTACCTIGIG **OCCUPATION** GTOCCCTOO OCATOCCOCT CCTATIGITICS **OCCITICATITICS** OCCOMPCCO CCCAGCGCGC TOCCCCACA TOCTIGITACTA CAGCCCAAAA CACCCACCC ACCTOCTCAC ACTICACCCA TOGETACAC **OCCTACATGF** GACCOCAGGT AGGACCATTC ANTOCOCTAC AATCCTCACG CCCCACCCC CAGCAACGCC CATGTCTCAG CATCANATCA AATCCOCTTC TCCCTCCCC COCATCACCA CCACGTCCCT ACACCTTCAA CCACCCCCCT GTCACTGCCG GACGACCTTIO **ACTUCTOSTIC** CCCACCACTA TOCCTATTOG TCCAGTGGTT CANTATGCAC CCCCCATGA TTCCCCATCC CAGGGTGGTG TGGTTGAGGT **AATAGGCTTIC AACCOSTICGIC** ACGTCCCCTCC GTATCCCCCA OCAOCCAGCC COCCOCCO OCCUCCOCCO CGAACCTOCC CCCTGCGCGC CAGTGTGGAA CTCCTTCCCC MOCGCCCTO **ACTOSTACCA** CATCCOOCAT CTCCCCCAAT TCACCCCCCA CACCCCCCAG TOTOCACACG TCAACAGCCC Gracicomo CCAACCOCCG TCACCACATT CACCOOCNOC ACCAGTOGTG OCTOCOCAGA OCCOCCOCACC CICCOCCCAC GAMCOCACO ATCCMOCACC CCCTOOCTCA ACTOOCGTOD OCCUTCACCO GTACCOGGGC **OCTTOCTICAL** TCACCTCCCA TCGTCCTTCC CACCCATICAG TACCCAOCAA TIGICATION CCCCACGATC CATCAGGCCG 00000000 GACCACTCCC **ACANTOCITIC** ACATACACCT COCCUTT COCCARCCA COCTOGATICA TCTTTTACGTA CCACCACATT **OCTIGATOCTIC** TGAGGCACCT ACCICACCCA CGATCACCAT CACTGACGCC GITCGCTCCC CCGTTCCCCA CCCCCCCCAC TCCCCCCTCC CCCCATCCTC TATTCCTGAC TCCCCCCC GROOCTATITI AGACTOCCOC MICTOCCC TOSTOGAÇÃO TOCACCCACC GICTCATCIT TCACGTACCC 7601 8301 5101 1089 6301 6601 6901 7001 1017 1017 1301 7501 9001 8101 <u>8</u> 5201 500 540 5501 5601 5701 1069 6001 7601 8201 8401 8501 8601 1078 9801 200 <u>e</u> <u>8</u> 6101 5201 6401 6501 6701 6801 101 7901 8901 9101 9201 9301 9401 9001

٠٤,

12600 13200 13300 13500 12300 12400 12800 12900 13000 13100 13600 10700 10900 11000 11100 11200 11300 11400 11500 11600 11700 11800 11900 12000 12100 12200 12500 12700 13400 13700 13800 13900 14000 14100 14200 14300 14400 ATCCCCCCAG ATCOTOTICG CGACGGCGC AAGGTACGTC AGAACAAACG GTGCCTCCTA CATTICATTIGT CAACAAATGG OCCCANOGAC CGACTOCCOS ACTTGACCCC CACAGGAGCG TACGCACCGG GICCCTCCAA CCCAACGACC ACGTACAAGG CCTTCGATTG GOTGAATTAT CAGAACTTCG CAGTCACGAA CGACGGGGGTG ATTTTCTTCT OCGAGGACCG TCCGATCCAC CCTCCCCCAA GATATCCATG AAGAACCAAT AAATGCGTTC **GCTCCGAACT** TCAACCTGCA **GCATCGACGA** TCATCGGGCA TTCAGTTCCA ACCCCCTCAC ATCITICITICIA CCCTIGITACITO GACCAACCAG GCCCCCTAAG CTCACCAACA CCCCCGATCT CCACCACCGA CACCGACTOG TYGAGTCTGG TCAGGCATCG TCGTCAGCAG AATITICGACG ACTICANACG **GCCACATACC** TCAACTCAGC **GCCGGGGAGCT GCAGGACCCT OCCURCINIT** OCCUTATATICS TOCTOGGATO **GCCTGGCCGT AAA**CCCTTC TOCTOCTOCC CTAGATCCCT COCACCAAGT TCCAGGATTC GCGGCTTGGG AATGTCAGAA **CCCCCTCCCC** CHOCCCCTCC CCGAGATCAC OCTGACATTG AACTGGCAT ATATTGGCTA CTCAAGCGCA **GCCCCCCCAC** CACCACCTTT CCCTGCGCGC receracte CTGCAAACTA CTCCCGGAGG GTACGCATTC ACCOGITITIOG AATCCATCGC GAACGCCGCC GAAACCCTAC CTTCCCAACG ATCATCGTCG TCCCACCCT ACCAGTOCCC ATGATCGACG caceroceae ergaaceebe OCCCAAGACC TACTOCGAGG ACCANTIGAA TATCACCTCG GICTTCCCCA TTGTGCAAGG ACCCUACTICCE STEACTICACE TOCAGEOCEG GCATCGACAG ACACCCTTCG TOCOGTCGTC GCCCGCGGA CGAGAGATA COCCGAGCAA CTGTTCAGTG TGAAGACGCT GACGTTTCTC GTTACCOGAC ACCOCOGTTC TOCCGACACG CTTGTCACCC TOSCGAGGIC TCAGGCAGIA GACATGCTG CTGTGCGGCG GCACGTAGTG COCTGATAAA GCCGTCGTGC ATGCTCGAGC CCATGATCAG TYCOCANTOC TANGAGACCO CTGACCOCOC TCCCCAGCCC OCCOCACCGA TOTCGGTCAG CARGACGTTG AAGAGAAACC ACTACCOCOG TGACGACGCG ACTITACTACC CCGACCAGAA CCCCCATCCA CTCCATGCTG GCCTAGACTC GATCCGCCTC COCCOCICIT CCCTCTCCC COTTCCTTCC **GTACCGATCC** PTCCCGTTGG CGCTGCGGTT CGACAACCTG CACCTGACCG AGICTOCTAC GACTACCOCT GCATCTITICT AGTGCGGGGT COCCTTOCCC COCCAACATT GTOGTICANAT OCCADODGEA AGACCGCCAG GTCATAGCTG GACCTCCCAT OCCCACCCCC TCGAGCACCC CGATGATCTT CTGTTTGACC COTCOCOCC ACCCGTGC ATCOTGCGA CCCTTCCACC CTGANGTINGS TATTICACCCC ACCCGANGG CAGCGACCCG ACCGGCCGCC CCCCCGACAG GTCCACCATC TATOXCTACC CGAAATCACT GCAATCAATT GGAACCGGTT TTGGCAAACG OCCCACCTIGA CCCACGCCGA CACGACTTICT ACGACGAGTT GOUCTGGACC COCCOGIATE TTECTAACTE CAGGIGAGGT TCACCCACGC CATTACCGCG TCCATCGCTT AACCGAACCT CAACCCTAGG CAAGGCATTC COTOCINGO ATACAGOOG ACCCCGTACT GITGGCGCGA TCTACTTGCT CACCCTAACG GCTTATICTICS ACTACTACCG AATCATTICGC GAGCGATGCC GOGNATACCC TOCTCGCTGC TCAACGACAG COCTITICTICC CGTTTGAGGT GTTCCCCACC TOCOCCCAN TONATCOCCC CTTOCACAAC TCATCOCTAA ATACCAATGC ACCCACCAAC ACCTGGCATC CCACCCTCCC accoacata CAGCCTGCCC CAGAACOCCG GCOGCACOCA ACACCGATCC OCTOCCAGTC COCCOCCAT AAGACCGGAT CACCETOCCC CAGCGACACG AAGAATAAAA **GCGGGTTGCAC** AAAGATTCCG CTCTCCAACG CTACACCTCG AGCGCAGATC ATCGTCGACT ACTOGREGATE ACANATOTICA ACGAGGITTIG TOCATOGICA GANGCTCATA CGTAATOGGC ATCTOCTAGE TOCOGNICTOS CCACOOCCOC GAACCCCCT ACCOUNTICA ACCOCCAGO ACCCCAAGAG CTCCCCATGT CCACTATOCG CAGCCTGCAT **COCCCC ATCG** CCCCCCTGCG ACATCAACAT CACTCCACGC GICCCACGIT CCCAGCTCTC TOCCATOCCC AGCCAAGACG TCGGTCATCT CCANCCACCA CACCOCCCCC TCACCTCCCC ACTODOCETTIC CCCCACAAGG CCTGGGCCCGA COCCCTCCCA ATGACCOCCA **GCTTGCCCGA** TOCAGOGAMA CCAACCACAT TCCCCCTCCG CCCCACACCG CCGCCCTACA TOCCACCTG OCTTOCACCC CCAGGTATTG TCCCTOCTAG CGATAATCCG TTTTCCCCCC 1GTCCTCCCCC GATCGCCATC TATGTCGGTG CAAGTGGGCC CCACCCAGCA CCTGAACAAC TACGCAACGG GATCTTCTCG **GCGGCCA1CG** TTTOGACCCC CTCATCCCAT GROSTICTICOS TROCOCTICOS ACCOCCCATG COCACTCTAG GANGAGAGCG TACCCCCAG ACCITCCCCTG COMMETEC ACCCCCCCCCA OCCUPANCE TACCCCATTIC COCCCCANAT COCCTTCTAC TACATCCCT CCTCCCATAC AATCCAATCT **acremental** CCACATCGTC CAAGGTCTAC TOTTOGGE TIGICOCGIT GNACCOCAGA ACACAACCOC COCATOTOTT CCTCACACCT TOCTCAGGGT TOCACCATICC CTGCCGCCAC CCAGCGCCC CCTCCTCCAC GACTCAGTCC OCANDOCAGG ACGOCICTICG COCTACCCAA CAATCACCTA TGACACGCTG ATCGATAGCG CCGAATGCCG TTTGGGGGG CGGAAGTGAG OCCUATOCCT TCATCGTGGT CATCOCITTIG CAGGCGACCG CCATACCCAA ATCCCTCTTT ACANGGAGGA **CCTANCETTT** CCCCCTTCAC **GOCCTTCATG** AACCCITACC TOTTCTACTC CCACCCCTAC OCTOCTOCAC ACOCTCTACG CCACTITITIC TCGAGCACCA GICGACTCAG ATATCGCGGC Crorcorrer **AATGTCCGAC** CTCAGCCCCC CCCTOCTOCT CCCCCCCAAC **GTTAGAGGAT** COTCATOCTC OCCATCAACT CTCTTCGGCA ACCGGGCGAT CTOCCCACGA CCCAGAGCTA TACCGATCGC GAMACTOCCC CONCECCO COCCUCAT CCAACCCCCA GGACCOCCAG TACCACCACC TOCOTOCICO OCTOTACCTC GACGATTTTC TOGTOCAGGT COCTCTCCA 1ccccrcrcs ACTTCCACCC ANGATOGCAC AGCOGGTGCC CACCOTCCAC COTCATCOCG ANGMANTCOC TICGATOCTG TCTATOGTOC ATTOCCOCCTIC ACCOTCAATO CCCACACCCT TTCCCAATTA COCTCACCTC OCMOCUTOR ACCITICACC COCCUCIA TGACCAACAT GTCCCAACAG CAGRECTITEC ATGAAGCCGT **OCCANCANGG** COCHEMICCA GTCCCCCCCCC CTOCTCACOG **OCCAGATTICA** CCACAMGTTC CCTACCCAGG CCCCAAGCAG TTACCTOCCA TTOOCCCC GAGTATGGCG CTCCTTCGGT TCATCATCGT TOTTGCGATC GTCATCACGT ATCACTTOOG TCCAGTCCCT TOCCOCTCT ACCCCCTC **ACACCCCADG** ACCAGCTOCA **OCCUPACTIC** CCCCTCATGT CCCTAACCGT CATGACCATG ATCATCCGAG ACTIGCOCCIGA CATOGETETEC CACTCOCTAG ATTCCGTTGG CTTCACOGTG CGATTACCAT OCCOCOCCCC CACCCCCTCC CCTTCCACTG ANTITACATOS COCAGACOCO GACAACAGGT CCACTTCCTC CGCTAGAGAA TCTCCACCGT ACCIOCCIOCAT COCTICITICOS TOGATCATCA CCACGTTCTC CCGCAACCTG COCCOCCTICGG TTCGTCCAGT TOTTACCCCT TOCTITICATE 900000000 **NOCACTAGGT** TOCCTICOCC GICCOTIGICG **COCCTATICCG** GTCCAGGTAC CCTCACCCA ATOCACACOG ACGTIGTICGAG **OCCATGMAD** ACCITCIOCCA GACGACGATC **GGTAATGCTG** ACACCTOCCA CCCTTACAAC CCCCCTCCAG COCCCAGCGC COCCATGCCC TATGTTTGTC TOCOCATICAL GATCITICATO **OCCCACCCGT** ACCCCCCCCT TATCCCAAGC COCCTATIGIT COCCATCAAC ATGTCCCCCC **COTOCOGTOS** TACCTACCAG ATCCTCATCA CCCCACACA CATECHEGE OCCADOCCTIC OCTOGTOCAC CCANCAGCGG CCCAGCAGCA CCCCCACCAA AGCATCGTC GATCCCCCTG CACCOCCAAT TOCCCTACAT ACCITCTICGAC CACCGATACC NGTCCGCGAT GCACCAGGCC **AACCAGACCG** TCAACCCCCC cococrocro MOCTOCOTT TCCCCCCAT CCCCTACACA THEOTICICAL **accracter** COCCOCCTC ATTICCCAACC ACCCCCCCCA GAACAAGATC TTCCCCATCA ACCACCTCCC GATCGCCCGA ACCCACAAGG ACTITCTCGAT TYCCCGGATC CCATCCCACC CCCACCCOCC COTCGATTCC GATCGTCACC ACAMGITTICA **OCCUPATION** COTOCITAGE CACCOCTITIC COCCCATOOO GTGAGCGTAC AAAGATTTOC ATCGACGATG GACTTCTAAC CAAGAATCGT 11701 11801 11901 12001 12601 12801 13101 3201 3301 13401 13601 13801 100 100 1101 11201 1301 1401 11601 12101 12201 12301 12501 12701 12901 3001 13701 3901 101 1201 1000 0201 1000 10401 10501 1090 10701 080 080 1001 11501 12401 13501 14501 14601 4701 9901

14901 TRATGIATOR TOCTAGATAC CTATAGGGG GOOGRAFOTO CTCTGATGGC TTGGGGGCG ACAGGGGGAT CACROSTCAA GCGGTGGCT CGAGTCAGGC 15000	15001 TORCHARCA COACATOR GCATCAACAG COCCCCCAC GOCCOSTAT COATGCCC ATCCTGACC CCTCGATTCG GOCCCCCCA CCAGACCTT 15100	15101 CACACCAC CHARTRACIC CCATGOTICS CANGOLACT TOCTOCTIAN CACCTOTIAN TITITACIDADE CCCCACTICE COTTUCTION CAGGITITICA 15200	15239	001 1 06 1 80 1 90 1
TOCTOCC	COSTGAT	TOCTANC		20
DATIGIC CTC	CCCGAC GOO	SCANCT TOC	WITC	0
AGCIGCE GOOCK	CANCING COCCO	DOTCOT COAM	OCCUPAC COCCU	20 30 40
AGTATC CTAT	GATCAG GCATA	PROCCA CCATA	ICATOT COTT	2
STATCT TOCT	MODGA CCAC	ACCOCC GAGG	CHOCKEN COOK	2
14901 TGATC	15001 TOOCK	15101 0000	15201 OCCICACICA COCCICATOR COTTOCOCAC OCCAATTIC	

Page 1 of 3

9/16

2500 2700 2800 2900 3000 3100 3200 1400 2100 2200 2300 3300 3400 1000 1200 1600 1700 COCOGCCACC CCAACTOGTC GAAAGCCTTC 1900 2000 ACATCGTCAA 2600 3500 3600 3700 3800 4000 1000 4200 4400 GACGCACCGA CATCACCGAC CTCACCCTGG 1300 TOCCCCCCA 1500 CACCTTCACC TOSCGACACC CCAGGCCCGG CTAGTCGCCC GCCTGAAGGG 1800 TOCAAGGCGT OCCACGGGGT GCCCATCCTG GCCGACAACA TCGAACCGAT 2400 OCTACCCCCA 1100 CTCTGAGGGT AGAGACGTGC CAGGAGGCGA 3900 GACCAACCAG CAATACGACG 700 GAGGIGITGA CCACCCCGGA ACGOCTGCGG TCTCTGGAAC 200 GAACTIGGGCG GCACGCTGTG 300 CTGATCGGCG AAGGCGCACA TCAAAGTGAT TCGCGCCCTT TTTCGGCCCA 500 OCCUBERCA 900 GOCAGGCTIGC GGCAAATGCG OCTOSTCAAA CAGGCCTACG CCGACATCCT CGCCGGGGCG TCCCTGGGCG AGTEGITISTE CAGEGGGTEG GCAAGAGEGG CAGGATATTE AATCECGAAC ACCTOCAMAC ACTIGITACIANG AACGTTICCOC GACTACTIGAC COGACTOGCC CONCOCCATG TOOCTICTICG GCACCCTICTIC CGAAAGCGAC ACGCCGAACA GETTICGACTE COCCENTITIG CTICTCCCCAA GCAGTCCCAC CTGTCGACCA GACCATCCCG AGTICATCGC COCGAAAACC GATCTCGTTG CCGCGAACAT CCCCACACCC TCACTCCCCG ACCITATION TCGAACATGT TAGCGAATAG CCGGGAGGAG ANCHICCACG GCCACACCGA ATGCCTACCA CCACCCCACC TCGACCACG CCCTOCACCA CTACAACCCT GTCGAAATTC TTGCGCAAAC CCCGCAACGC CGATTOTICOS CAAGOCCCAG TOSTCOCCOC TOSTOGACGA GOCGACOTTC TOTTOCOCCO OCAGATTICAC GACGCCCCCG AAGCCGAAAC CATCCCCTG GACTOTTICGA COCCATACCO TTCCCAACAC CCCAAGTCCC CCCCATGATA GICACCACOG CCOCAAAGGC GATGCCCCCA TCGAACACGA GCGCTGCGAT GOCOCCAGCG GTGCCAACAT TACTCCACCA CCACAACCAG CCGCGAGCCG GACOGGIOGG TOCOOCTGCA CTOCCAATTIC GCCTGCCTGA OCTACGGCGC ACCACACACA COGOSTICACO GACCTOCTICO COCOCOSTICACIÓN STYCOCCACAC CGTCTAGCGC OCCCCAGCGG ATCATCCTGT TCGCCAACGA CCGCGGCTGC CAACCACCTC ACCCCATCAA CCGCCAAGCG CCGCCCCCCC TCCATCGCCG TCCCATCGAG CTGGAAGCCT ANACCGACCC CCAATTACGC GACCTCCACA CTCAAATTOC GOCCCCACGC GCCCGCTGAC CGCATCACAA AGATATGCCG GOCGCGACCA ACCCCGACGA **GOCGCCGGCA GOOCGGGGGT** CCTCCCACCC TCACCGGCGA COCACCAGCT GCCCCCCTAC GAAGGCTAC TGACCCCCC CCAGGTGAAG ATCAGCACCC GCAACCGGGA CTCCCCCCAA CAGCTCCCCC GOCCGATGGC TCAACCACAA GTCCCAGCAC COCAACCACG ACGGGCTGCT CCTCCAAACC TACATCCCGC CCACCCTOCC COCTOCCACT GACGCTAGCC AACCAACTOG ACACCCAAGC CAGCGAGGAA CCATCOCCGA COCCOCCGAT CTCGGACCTC GACCTICOCCG CAAACCCCCTC AAATATCGTC CCATCACCCT TAGGENARY GOCGGCCCCC CTCGATCGTG GTCACCACCA CCCTGACCGA CTACTCCCCC TCACCGGGCT TODOCCACOG AGGACCOGGC ACCCGCGACA **CCCCCTCCT OCCOOCCUT** GATCGACCOO GACOGOCACO TOCTCATCGA COCCCCCATC CAGACCTOCG CCAACCCCTC TCACACACCC CATTCCCCCC CACCOCCTOG ACCAGCACCO CAGAAGACCG GAAACCTGAA AAGCGTGCGC CGCCACCTGC CCACCCCACC CCCCACTGCG OCCCOCCAGA TOGTOCCACG TOCTOCACGE COCLARACCOC ACACCCCCAG CCAAGCCCAA ATCACCAGCC ACGCCCACCA CCTGGCCGCC GCACTGGTAC CTCCACCGGC CCTCCATCCT CCCCCCCCCA CONCINENT ACTACCTCGA OCCTOOCAC CAGGACAAGC CGAACGCCCC GAACCCGTGC TACCTCCCC CTCCCAATCT TOCOCOCOT GAATTCACTT ACCTAACACC ACTTCTAGCA GCTGTCGGCG CGACTTCTTG ANGOCTOGAC CACCCOCAAA OCCUMANTOS GROCCOSTOS COCCCOCTG GACCGCGACG OCTAGACTICE GACGTAGCCG OCTOGATOCC GAOCTOGACC CCTTGGACGA STETIOGRAFIO CITIOSTIGGE COSTIACCOS COSTCOSOCA CACCITIGATO COCOGNOGAT GIGTCCAAAC CCOCCAGGCC GCCGAAGCCC CACAGCCAAG CCCACCACGT CIGIGGEORGE TOCTIGICTAC **COCAMONICO** COCCOSTOGI CCCTOOCCAC CCTCOCCOOC CANDOCACCA CAGCGCCGCG CCCCACCCC GGACAGCGCC COTTCACCAT CACCGGCCGC **GOCCOCCCC** CCGCCCCCAA ACCOCCAGGT COTCTACGTG OCCIPOCOCT CGAACGCGCC GRECTEGACG TECTEGACIGA CCCAACCCCC TOTAL COCCOCATO CTCACTCACC COCCOCCGAC GCGACGATGC GTOCICOGTG CACGAGCATG COCAMIDOS COCCOATOTO GACTATCCCG ACCCCCCTG CITICATCAAC GICCCGAATG GCACCCTCCA CCTCCACACG CTOCOCCTO OCCANOCOGY TACOCATCAC CANOCOCOGAC GCCGCCCTAC CCACCACCCA ACACCAGGGC CACCCCACC TCACCCACAC CCCAACCCC GCCCACCTIT STCATCCACA CCACCCCCCA TOCOOCCOCC ATCCACCOCG CLACACAACG GTCTTCCCGT CCATCCCCCA TGTGATCCGC GACCACCOCA CACCCCTGCC OCTGTATICAC ACCAAACGCC CATOCCCGAC GCCGTTCACT COCCANGATA ACGARACTTC ACOCCCOCCA ACACCATICAG GAACOCCTCC CCACAACATC GTCATCGTCT CCTTACCCCA TACGACCTCA AATGCGCCGC GTOCAGTICOG ACOCOGTOGT CCCCCCGACTG GCCCGGCATC ACCOGCTAGE CGAACGTTAG ACCUANCE CCASTOCTOS accordance econocinal ACCOCCCCC CGTCCCCCTT ACCCAGGGCT OCCOCCCAAC GAGTETCGCC CCACAGITIG ACCOCCACCO TYCCCTYCYGG GGAACTYGGGC OCTOTICACOC ACCOCCTAC CTCCCCCAAA TOTOCOOCCA ACCCGACGAG AACCGCCTGG **GCTCACGAAA CCGAGCACAA** CCCCCCCAA CCCTCCCCAA CCAGTOGAAC GACGCCGGGG OTOCCCCCTA ADDCCOSTOCT GGACGCCCCC TACCOCACCG CCTGTATCAC ATCGTGGCCG AGCGGCTGGC GACCTCCACA GAATTOCCCA TGAGCCGGCA CCTCCACCAC COCETITIOCC COTOGTAGCT GTAGCCGCAC CCCTCCCCCC CTACCOCACC CCANTCCCCT OCAGOGGACG ACCOCOGGA COCCOCCOC ACCCTOCTAC TACACTACAC TOTTOCCCGA CATCACCGCC GFTCCCCCCC COCACOTOCG ACCAGCOGCC **accessors** CCCTCTCCTC COCCGATCCT GTATCTGCGG GAMANTICCEA ATCOCCOTCC TCATGGACTG OCTACACCCC OCNIGICACO OCTANGIGOC TACCIGACCC CITICICCAGO TCITICATOC CCACAACCGA ACCTIGTICAGA TOCCTANACC OCATATAAGG GOCCGGCAGC CCCTCTACGG CGCCCGACCG CCATCATTTG OCCUMENCE ACTACGCACC CACCCCAACG TCGCGGCCCCC CTCATCCGCA GACCOCCOCC OCCUCACOCO ACCTCCACCA **GTCACGTGAC** COCCCANGCC CCCACATCCC ACCOCTAGCC ACCTICOCCCOC OCCUATOTIC ACCAMACCCG CCTGCGACCC CCAACCCTGG ACCCCAGGAC TACCAGCAGA COCTACCTCC ACCITCICCCG COCACTACOC TOCOCCOCCC TCTACCATICA **OCTOCCCGAT** accraacerc CACACCCTOG **GCGACCCCGCA** OCCCAGCGGG CCACACCCCC COCCOCCTCA ACCCTICAC TCATGTCATT **OCAACCACCT** GAACTOGAAA TTATTGGCTT COCCOOCTIG MACTACGCGC COCCOCACTAC CGATCCTACC TOCCOCAGOC COCCICATOS AAGTTCATOG OCCUPACION GICGGIGGCC **GCCCAOCTGT** COSTOCAGGT CTTGTCGCGC 1501 2301 3201 3401 1201 1301 1401 1601 1701 1801 1901 2001 2101 2201 2401 2501 2601 2701 2801 2901 3001 3101 3301 3501 3601 3701 3801 3901 1001 101 1021 2 30 5 50 5 107 801 1101 901 9

Figure 3 continued, page 2 of 3

5600 5700 5800 5900 0009 6100 6200 6300 6400 6500 0099 6700 6800 0069 7000 7100 7200 7300 7400 7500 7600 7700 7800 8000 8100 8200 8300 8400 8500 8600 8700 GCCTTATTCA OCCOCTCACG 7900 8800 8900 9000 9200 9100 9300 ATTCCCTATG TOCCATOCCT **OCACCTATICG** ACCOMMAND CACCCCACTG CAAAGCGTTG TTCCCCCCCC GCGAACCCGT CATGOGTGTC TIGICTOCIC GCAGACOCCT GAAGCCCGCAG ACACCATCCT CCACTCCCAC TOCCOCCTCC CCGGCGCAC CCTACGACTC ACCTICAATGC CCACAGTACC GACGACACCA CGACCTATCG AGCAGGACGA GTTCCCCCCC **AAATTICATOC** CTCACAATTA GCCACAAGCG ATTICCGCGCG ATTCGGGGC ACAAATCGCC GGGGTGCGCG CTATGACCGA ATTCGACGAC ATCAAAAACC ACCOSOCCOS TEACCTEAGG STOSTECCCS ACATCCCTTT CCGAGTCGCC CCTTICACCTIC TTCCCCTCAC **GTCTCATACA** COCACCATC CACCOCCC TOCOCCUIT CCCCCACCGT TOTOTOCCAG ACCOCCTGAT Menester interested interested interested COCCACCCOC OCTOCOGGG GTCATCCCTC GTACTCCGAA CTCCACCCTC **GTTCCACTCC** OCCCTCAAAT CCACCAATCA TODOSCICAN TOSCICACIO GGAGACTACC CCCTACKTAG Tata de para la NEGREDARM INDIRING MORRAGEC ANCCAGAN CCAGCGGCG CATTIGITIES **ACTIGOCOCOTA** ACCOCCCACCC CONTROPORA OCCAACIGGAA AATCTIGTGTT CGACAAGGCG CTCCCCCCCCC GCCATCCCCG CCCCAAAGAT ACCCITICCAG TTCGAGGCCG ATTACCCCC CTACCCCCCC ACCOCCCCCA GTCCCAACC CGTGATGGTC GTCCCATTICG CTGCAACAGC CTTTTCGGG CTCCATCGCC CCAATCCCCG COCCITACCITC CCTAGACTAA COCTOCCCA ATOCCOCCC GTCOCACCA CTTTGGACCG CCCCGAAGCC **GCTGCGCTTG GCCTCCGAGT** GTTTGCCCCG GACCCCTGAT GACCTTTTGT OCCUANDEC GAACCETOCA COCAMAGICC GCAGCAGCCT CACGCCTTGT TOGGIOCOTT TOCGACCCAT CAGCGCTOGC TOGCGGCCAC CTCCTCTCCC GGATTCCGTT CTCCATCGAG AGACCOCCOC GITCGTCAAC GGCTCCGGCA CTCCCCTCAC **AACCTCCTTG** COTTCATTICG GCGGTGACAG CCATTCGTN ATCTTTACGC ACCCCCANAC TTCCANTGCC CATOCCCAOC **GGTGTGCGCG** ACCCAACAGG CCAAATCCAC CACCCCGACG CCCCAACCGG ATGCCCGATA GCAACCTCCA TGTTCGAGGC TCACAACCC CCAGGTTCGC CTOCCOTTC TOSTITICATOS ACTICCCCCCA CTGCAGACCG ACCONCINCO CCACTCCAGG AACACTTATC CTCCGCGCG ACACCGAGAA GCGAGACGCC ACCCTCATTA GOOCOGITIC ACCOCCIECG **OCCOCCATOC** TCCCCACACC CCCACTOCCC TTCCCCCCC ACCCCCCCCA ACTOGATTICT CCCONCGATC CAAGCCCCCG GTGCCGCGG AAGCCCCCCC CCTCACCTGT COCTIONAGE TOCCGGCGGA OCCUPACING TACHGAAGGT GITCCCITICIT CATCCOOCOG CCCGCGTTCT GCCCGATGAA CADGACCOGG AACTOGACGC CONCINCACO TOCCOCCAC ACTACCCATC ACCCOGGICAC TCACCCATTT GCCCAGGATG CGGTACCTAC CATCCOCCC CACCCCCTCC GAMACCCCCG ACCACGCCGA CCTGCCGAAC GTGCCCGTTC CTAATCGCCG OCACCAOCTC CACGGTGCCG GCGCGCATGC CACCAGTOCA TATCCCOCCA ACGACCCOCT **GCATGCATCT** COCCOCTCCC CCCCCCCCCA TTCAGTTGCG CCCCATAGCG **ACTIONOSICT ACTICATOCCG** CGACTCGTTT OCTCAACAAT CTOCOCCCT CTCGTGGCC ACCCANCOGG COCCOCTOCA COTCACTTCC GTTCCCCCCT CTCACCCCTC COTTTCCCAC **GCAAGCTGCT** ACAGCCATT ACCTTCCCCC ACATOGACAC **CCACTTOCTC** TCCCCTCCCC CCACCTCGC CCCCCCCCAA GACAACTGAT COCTGTAGGT GAMACOCTC MODECTICAC AGCCGCACGC GAMACCACT ACCCACCCAT **GTGACGCATG** ATCACCCCCC OCTOCOCCA CGTACATCAT COTCTACCTC COCTCAACCC COCCTOCAAC CTOCHEREST CCAGCAGCAG COCCACCC COCAGICCAC GAGACOCCCC CCATCCCCCA OCAMATGTTC CTACTCCATT **OCCUPICAC** CACGGTTTTGG 200000000 CCTAACCGCC TTTOCCOCC GCCCCCCCTC 1998 PR 1989 National States CTGATCCCCC GAACTACCCT GACCAACTOC CACCATCAAC GAAGTCTCCC **GCTCCATGGT** DOTOCOCINC COCCOCCANC TCGFFTCACC TCCCCCTT TCTCTCCCA CCGTTCTTCA CATGICCTOG CGATTCTTAT OCTOOCCGAA TCAACGATAA TAGTGATTCC TOCCGACGAG CTGACCTITIT CATGCACCGG **GCACTATCCC** COCCOGNOCT GTCCCTCCAC CACACCTAAC GROCCOMCT COCTIGECOCG GACCCCCCCA CAMAGICCIG ACCEPTORING ANAGACACA ACCEATOCET CATCGTGACT COCTTOCCTT CCACGCGTCG OCTUCTIGATE GOSTACTICES AGACCTICATE CCCCCTCCT GACCCCCTCC TCCCTAAGCC AACCAACCAAC CCANACACCG COCACACCCC ACTOSTICAG GACCCCCAGG GACCOFFICCG GACCICCCACC **GCGAGTTCCA** CCCCCCCCCCC ATCCTTTTCC CTACCCOGTG **AACCTAAAAG** TACACCITIC CTCCCTCTCC CGACAGICIC CCCCCCAAGC CAACTCGTTG ACCOSOCCOC TOCANTOCTIG TTGCGGGGCA ACCCCCCTTT ACCTCCGAGA GICCICCICC TCCCCCAACA GOCGGATTTC TITGTCGTCG CTGCGGGTTG CGTTGTCGAT COCCOMOCC OCACACCCTT CGAACTTGTC GICTOTOCOTO GACAGAGTTG Paradana sana sananana Macconstr issesses CACATIONOCT COCCOCCAC TCACCGAATC STCACCAGCT CONTRODUCA CCATCATTAG ACCOCCTGCC **OCCUCATICAC** CCCCCCCTAT ACCETCACCC CCCCTCCACC CGATCAACCG GCCCCTGACC CGAAGCACTIC GAATCCTTCG CATCGTGACC AGGCCCAGCC CCGMACCOG CCCGGCTCAA GOCCTITATIC CGACCTGCAA ACGCTGTGGT COGNIDECCE DEACTOOCTE GTGCACCGGA ACCCGGCGCT GACCICCION COCCOCCOC **OCCODOCCAC GCCGGGCAGT** TITICAGINOC CACTTCCACC CTCCAACCCG COCTITIOCCC **accecant** ACCACCTCCT COCCOCCOCC CCOCCIGARAC CITICIFOCCOC ATGGGCCTGG TGGACACCCA CAGCCGAAGC **GCTCCAA**CG TTGGCCCACC AGATTICTICGC CTCCCATCAG **CCGTTCCCCC** TACCCOCCAA COTCATCOTC ACCOGNICCC CCTGTTGCCG ACCORDICACAT CAGTECCACA CCGAATCGCC TACTAGGCAC COTCCGCGAA TTATOCCTOC ACCOCTGAGE CCACCCACTA MGATCGCCG CCGTTTCAAG CCACCAATGC TCCCCCAAAG CCCATCATTT CACATOCOTC CACCOGCOCT AGTCCTCGAT CCACCOCCAC **OCCUPITACA** CCCCAACTGT TOCCOOLOCC ATCCGCTGAC STOCCCAAT TCTCACCACC CCCCCTCCAA GACCCCCCCCA CCCCCCCCC CACCGTCCCC ACCIOCCACCA TCCCGTTCCA CAACCCCTT TTCCCCACC TOCCGACCCT CACATOGOCA TANCCOOCAC COCCACCCC CONCOUNT Netherly Netherly INDEPENDENT PROPERTIES NECESSORY TCCCCCCTT CCCCCCATOC COCCHOOFIC OCCIPION CACACCCAAA TCCAGTOGTO GTCCCCCAT **GCTTCATCGA** ACCCGAMATC AATCCAACCC GTGAGAGGG MACCAGCIAIA TOGITICATICA ACCTGCCCTA **STCOSTATOS GCCCANACGT** OCCIONCACCC **OCACTTACCA** CCAGTGACCC CGACCAGGG GCGGGGTCGC OCCANGACCA ATOCCOCCOCT TOTOCATOCO CCTCACCCAC ATCCCCTCTC **OCATICTICOCC** TOCCCCCCA GCACCCAGCG **ACCATCOCCC** GACCACCAC CTOCTAGGTG ACCAGOCCAC **CONCINCIA** GACACCCCCC ATGTGCTGGT TCAMACCOST AGGAACTCCC TACAGICCIT COCCOCCCAC ATTCGGCAGG TOTOCOLACGA GCCCCCCAGA **OCCUPATIONS** CTICGTICCC GTCAGCCCCCC CARCETTCCA CCACCACAGT CTTOCCCACT **GCGGATTICITY** 000000000 OCCOCCATOG COCTOCCTAC CCCTTOCCCA STOCTOGGGG **OOCCITCOOCC** ACCOUNTEC TYCCTOCAAT ACCATANACA ANGANGOCGT COCAATAGCG COCTTATACC TOTACCOCCA TOCCATOGIG CTOCCCOCCO TOCTICACCCG TCGCACCCCA TACAGGTCCA CCCCAGAACC CARCCCCAT OCCCACANCA **CCANTIGITICS** CACCACCCAT GAACTITATE CTCANACCTG CCTCACCGAA 200000000 ATCCAGTTTG TITCHIOCGE ACMATTICAN CCACATTOCC COCCANOCOC COCANTCOCT TCTCTTTACC COCCACCCC GTGCCCTACA CCCCTCCCAC GTCCTCCCCC CACTATTICC GACCIAACCCG **OCCCADGAGG** ACCCCCCCCAG **OCCUPANCE** TOCCATOCOT recherteering ACCCCCCACC GATAGGGCCA ATCCACCTG CACCOOCTIC 5001 5101 1807 5201 5301 2401 1901 5601 5701 550 5801 5001 6101 6301 1001 7601 5901 6201 5 6501 6701 5901 100 1201 301 1701 1018 8201 6601 5801 ₹ 1501 7801 7901 8001 8301 1018 3501 8601 101 9801 3901 000 9101 9201 5 500

	-	-	_	1 02 1 09 1	09	- 20	-	<u>۾</u>	07	₹ -	
12412							,	•	۲	TACACAT	101
12400	OCCOPORTION	AAGACGCCCA G	CICOCCIOC 1	ATCCCCACCC	ACCOCCTOCG	GTOCTTCACC	GTTCACCAGC	ATCCCCCCAA	CACACCOCA	GCACCACAGC	
12300	accendraces	CGACACCATC G	CCAGGATCAC (TOCOGCTOCC	CACTOCCOAC	CCCCCCCCCC	MACAGOCCO			CCCANGCAGC	107
12200	CCTCCCCCCCC	CCCCCCCCCAC	COCCACGTAC (CACCCCCACC	COCOCOCOCO	TTCACCCATG	creaceace	CCAOCTCCAO	ACCCGCCGG	ATCUSICAGE	1017
12100	COCCOACAAC	COCCOCCCCA	OCATICETICAL (GTCACCACCA	OCCOGNICCCG	CCACCCCCCT	ACCONCINC	COCCOCACACC	ACCCCAGCGC	COLOCOCCCG	1007
12000	OCCOOCCTGA	CCACGTCGAT C	TTGCACACCG (CTGAACGCC	COCTOCCOCT	TCACCOCOCO	CACCITCCTCC	CCACCTCCCC	CCCMCCCCC	CONCACCCC	1061
11900	COCCCACCTG	COCOCCAAGC C	CTGCGGATAT (COCCCATICGG	000000000	GIGTTCGGCG	CCATCCCCCC	CCCCCCACC	CTGATCGCGG	OCCOCACCAT	1001
11800	TCTCCCATCA	ACCOSTCCAGG 1	ACCOCCCOGG 1	ACCACCCTCA	COCCCCCTCG	OCHOCCCOCC	AGTTCCACCA	COOCTICOCO	OCCITCACOCC	ACATCOCOCA	10/1
11700	GPCGACGGCG	COCCOCCAC	ACCANAGOCOG (COTCACCACC	CCAGGTCCCC	TTCACOCTCC	critocicnoc	CCARCETCIA	OCMOCOCIT	CIGHTGIOCA OCANGCOCT	1091
11600	CACATGAAAC	ACCAGCCCTG (ACTOCCOATC 1	COTCCOCCA	MCCCCCCCC	GOCTGCCACC	ACCCAPTICGA	OCCAGOGOGG	AGCAATGCGC	CCCCCCCCT	ומכו
11500	OCCAGAGCGG	000000000	CCAGCGACGC (TCCCCCCAT	COCCOCCAAG	CCCCCCTCAT		CCAGTTGCGG	OCCCOOCNAC	CACCCACCA	160
11400	AATCCACCAT	ACCTCGATCG A	GTCGAAGACC	COCTATCCCT	ATCCCCCCT	CATCCCCCCG		OCOOCOOCCA	OCCTATGCCG	OCOCCIOOCT	130
11300	pochocco	CACGTGCGGA	COCCAGGGA CACGTGCOGA COCCTGGCCG		COCCCATC	CATCGACCTC	CCCACCCCCT	GCCMCCCA	OCHOCCOACO	CAGTECAGCG	1201
11200	GCCCATCACG				COCTGAACAC	ACCACCAGCC	actidaticac	CCAAAAGCTC	OCATOCAGGG	1 STOCACCTICG OCATOCAGOS O	1101
11100	COCAGACAC	Acceptance (COCCOCCCCC		COCACOCTOT	CCACTGACCG	CATGGGCGGT	OCCAGOGIGA	CACCCCTCCC	OCTGACCACC	100
11000	COCCCCAATT		GTCCCGCAAT		ACCGTCAATC	COCCGAACCG	COCCOCANOC	OCCURACION	CACCOCOTICAT	NOCTACTOOD	901
10900	TCACCCCAGC		TCAACCTTT	CCCTTCCCCA	OCCATCANC	ACACGAAGIT	CTOCCCATTA	AACCCOCCAAO	CACCCCCCCC	ACCTEGACING CACCEGACES A	000
10800	CONCINCOCC	CATCCACCTO	CACATCCCCA	AAGATCTTTG	000000000	COTGAGAATC	CACCACCATC	CCHOCHCANC	OCCIONOCIA	MACTOCTAC	0701
10700	CCACCACCAG	ACACCTTCCA (CTTCAOCAGE	GTGGAATGTT		GICATOGITIO	TOOTGTGCTG	TCANACCTOC	1090
10600	ACCAAGTOCT	OCTTACCOTC 1	TPICCAGGIC	COCTCACCOC	CCACACCCCC	GETCGCGGTC	TODOCCACCT	CCGCATCGTC	CONTINUE	COCCCCTCCC	0501
10500	CACCTCCAAG	SCITTCET (COCCOCCTIOC	OCCGATTICAA	GIGICCICG	CACCTCCCCC	OCAACOCTTC	CTCATCCCCA	GTOCTCCACO	9401
10400	COCCERCOCC	Tracractics (CCCCACTCG	OGATCCCCGG	COCCOCCCA	COCCCOCAG	COCCTIGING	COCTICATOO	TCCGAGTTOC	COCCOOCCAD	0301
10300	xeregreer	GICOCCOTICG COCTICGICGE	COCCTOCACG GCGCGCTGTC	COCCTCCACG	CCCGATCATC	COCCOCCANG	acroaacri c	CICCICTICC	ACCETCAGCG	COCOGNOTIC	020
10200	GICCIGING	ACCCCCAGC (COCCCCTCGG	ATCCATOCCC	OCCOGNACCE	COCCCATOG	AGATAGETTE	TOTTHETECO	COCCOCCATT	COCCOCCOC	010
10100	CTCCCCCATC	GCCCCTCCAG (OCCCCTCOCG	CCATGATCCT	ACCTOCCO	CHGANTATOD	COCCCCANC	GTACCCATO COCCCCAA	000
10000	ATTICHOGITIT	TCACCCTCTG /	TTTCCTCCAA	CAGIGICIGI	OCCCCCATAG	OCCCAGGGCT	COCCOCACAT	COSTAGOTOC	TITICACCC	TOTOGRAFICO	9901
0066	TOGGTTOCGT	GICCICCCC 1	OCATTICENT	TIGITITCIEC OCGANCOOC	THOTTHCTCC	GCCCCCAT	CCOOCGACCA GCOCCOCCAT	OCTOBOOGNT .	OCCOPTC	CTICOCTITIC	9801
0086	STROCCOCTC	CCACCCTCCC	COGGENITIOS TICKTOCATICS A CUAGGEGA COACCOTOGO GITTOCOCTO	TCATCCATC:	CCCCCATTOC	COCACCARAC	TCACACGCT TTCCTTCCAT COCACCAAG	TCACACOCT	MOTOTOCOCOC	970) cercerrore Acrerocese	9701

					_	_ 0			•				
		<i>WIIIIII.</i> BCG∆ = ~8.8kb	150	Homologue Accession #		246257	U01072	X79562				L29506	A08331
	æ —	+ 9		P value		1.48-14	3.0e-13	2.38-43				3.6e-16	4.86-14
	n Nh St	12 13 14 15	1F 1G 1H	Homologies to Predicted Encoded Protein		M.leprae aceA	BCG uraA	M. luberculosis esat6				B. subtilis subtilisin	Serine protesses
	N -		<u> </u>	Encoded Protein	57 57	36	70	90	54	48	50	46	-
	xb St — Xmillimillimillimillimillimillimillimil	- 60 - 80 - 1	10	Possible Ribosome Binding Sites	AGGA (10)	GGA (4)	GAGG (5)	C DDVD	GGA (5)	AGGA (9)	GGA (11)	GAA (5)	
	9x ————————————————————————————————————	1° 1	51	Start - Stop (base pairs)	889 - 2433	3130 - 4203	5075 - 6046	6954 - 8612	10619 - 9663	13328 - 11946	14823 - 13438 14643 - 13438 14541 - 13438		
Q	I Нр	4	18	ORF Size (base pairs)	1542	1071	696	1657	954	1380	1386	1368	
Region 1 (16.9 kb	- H	~ 1	A1	M. tuberculosis Codon Usage	yes	yes	yes	yes	yes	yes	yes	yes	
Bec	፫ ─■-	- 0		ORF	4	18	ပ္	₽	<u>"</u>	=	1G	Ξ	

Figure 4

Region 2 (15.3 kb)

	<i>'''''''''</i> BCG∆ = ~10.8 kb QRF	Homologue Accession #				F24194	U00015 A00975 U03393		X73226	A30545	X65104	Z22594 VE0166	20.00			
≅ ─ 1 ↓	90 90	P value			0.05.47	9.9e-47 <1e-5	1.5e-7 -4.e-5		9.9e-146	6.7e-141	3.1e-11	4.46-11				
2 - 1 + 1	E	Predicted Protein			Vi2	Y Alim	B1620 ORF ses		n RNDPR	is mol64	ретеазе	ndal gene				
×	1 12 1 2G ♣ 3	Homologies to Predicted Encoded Protein			Airi doo A	L.com con lysR family	Mieprae cosmid B1620 ORF Cutinases		S. typhimurium RNDPR	M tuberculosis mp164	E. coli gabP permease	T. harzianum indal gene retroviral recentor				
+		Encoded Protein max. ~kDa)	25	16	34	34	22 M	19	37	24	·		31	35	36	212
_	20 2J 25 2E		AGGGAG (7)	AGAA (4)	AG2 (B)	GGA (8)	попе	none	AGGA (11)	AAGA (6)	AG (10)		GGAAGA (6)	GAG (10) GGAA (8)	00A (9)	AG (10)
E	SK ↑ 2C ↑ SK		+	2862 - 3298	3003 - 3590 5187 - 6134	5376 - 6134	6561 - 7217	8036 - 8560	9941 - 10909	11118 - 11783	11965-13407		14221-13376 G	8259 - 7211 7939 - 7211	4992 - 4327	5117 - 4521
	SA 282 - 281	ORF Size (base pairs)	558	43/	990	948	657	522	996	999	1443		846	1050	999	597
w w Bg	-	M. tuberculosis Codon Usage	yes	yes	Yes	yes	yes	yes		yes	yes		yes	yes	yes	yes
<u> </u>	-		47 g	283	3 8	3	2D	똤	2F	52	75 H		2	2	×	ಸ

Figure 5

_
-0
×
4
2
Ξ
n
0
0
0
Щ

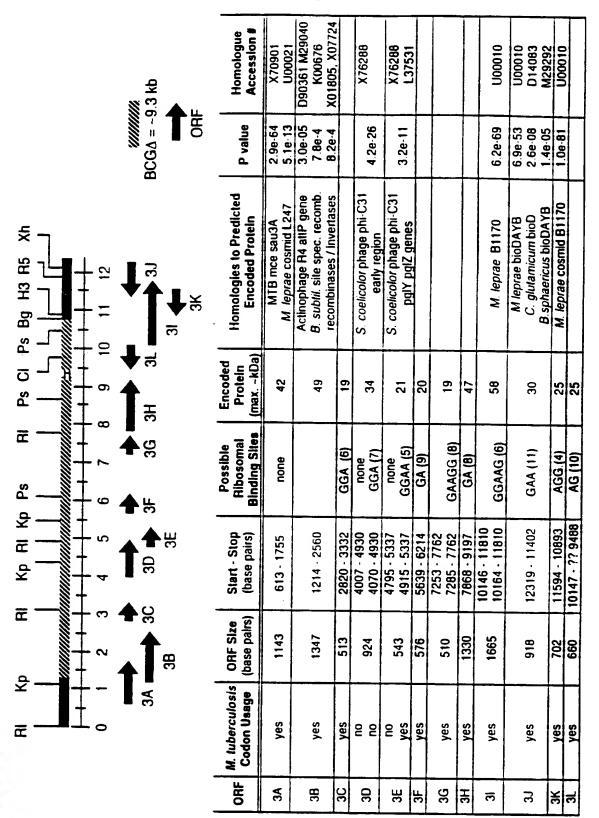


Figure 6

Region 1

1 1	
CCAGCCGCCCGGATCCAGCA	A BCGA16 Junction
CACAT GCAGCCGTGGGTGCCGCCGCGGGTGTCTTCATCGGCTTCCAC CCAGCCGCCGGATCCAGCA	
rGGGTGCCGCCGG	BCG∆1
CTGGTCGACGATTGGCACAT GCAGCCGTGCCGCCGCGGGTGTCTTCATCGGCTTCCAC CCAGCCGCCCGGATCCAGCA	BCGA1a Junction

Region 2

CGATGATCTTCTGTTTGACC	BCGA2b Junction
CCACCC GCGCCCCCGCTCGCACTAGA	BCG∆2 BCG∆2
CAACTCCACGGCGACCACCC GCGCCCCGCT	BCGA2a Junction

Region 3

-----CACCTCGACCACGGCCAACC|GTGGACCTGTGAGATACACT------TCAGCAGTCCACGGCCAACCC|CCGCAACACCTTCCACC----BCGA3b Junction ---- -9.3 kb BCG∆3 BCGA3a Junction

Figure 7

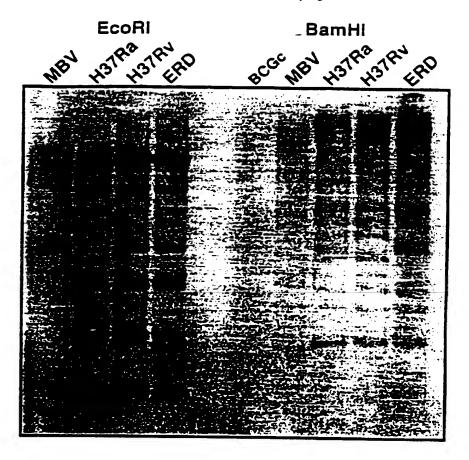


Figure 8

International application No. PCT/US96/01938

A. CL	ASSIFICATION OF SUBJECT MATTER					
IPC(6)	:Please See Extra Sheet.					
	:Please See Extra Sheet.					
According	to International Patent Classification (IPC) or to bo	th national classification and IPC				
	LDS SEARCHED					
Minimum	documentation searched (classification system follow	ved by classification symbols)				
U.S. :			24.32, 24.33			
Dogumenta						
Documenta	tion searched other than minimum documentation to	the extent that such documents are included	in the fields searched			
Electronic	data base consulted during the international search (page of data base and subary an windle				
Please S	ee Extra Sheet.	name of data base and, where practicable	, search terms used)			
	as the shock.					
			:			
C. DOC	CUMENTS CONSIDERED TO BE RELEVANT	·				
Category*	Citation of document, with indication, where	appropriate, of the relevant passages	Relevant to claim No.			
X	Infection and Immunity, Volume	e 61 No 5 issued May	1 10 16 17			
	1993, H. Li et al, "Evidence for a	bsence of the MPR64 gane	1-10, 16, 17, 24, 25			
Y	in some substrains of Mycobac	terium hovis BCG" nages				
	1730-1734, see entire document	. Pages	18-23			
			10-25			
X	JP, 1-247094 (AJINOMOTO ET	AL) 02 October 1989, see	1-7			
	entire document.		• •			
		•				
X	Infection and Immunity, Volume 5	59, No. 10, issued October	1-17			
1991, C. Parra et al, "Isolation, characterization and						
1	molecular cloning of a specific m	vcobacterium tuberculosis				
	antigen gene: identification of a s	pecies-specific : rouence".				
-	pages 3411-3417, see entire doc	ument.				
}						
į						
X Furthe	er documents are listed in the continuation of Box (See patent family annex.				
Spe	cial categories of cited documents:	"T" later document published after the inter	mational filing date or priorny			
A does to b	ument defining the general state of the art which is not considered e of particular relevance	date and not in conflict with the applica principle or theory underlying the inve	tion but cited to understand the			
	ier document published on or after the international filing date	"X" document of particular relevance; the	claimed invention cannot be			
L* docu	ament which may throw doubts on priority claim(s) or which is	considered novel or cannot be consider when the document is taken alone	ed to involve an inventive step			
CITED	to establish the publication date of another citation or other ial reason (as specified)	"Y" document of particular relevance; the	claimed invention cannot be			
O* docu	ment referring to an oral disclosure, use, exhibition or other	considered to involve an inventive combined with one or more other such	step when the document is documents, such combination			
P* docu	ament published prior to the international filing date but later than	being obvious to a person skilled in the *&* document member of the same patent f	ent .			
<u>-</u>	ctual completion of the international search	The second secon				
u	order completion of the international scarch	Date of mailing of the international sear	ch report			
17 APRIL	1996	29 MAY 1996				
ame and ma	niling address of the ISA/US	Authorized officer	ENUL 10			
Commissione Box PCT	er of Patents and Trademarks		Fruit /0			
Washington,		JEFFREY FREDMAN	•			
acsimile No		Telephone No. (703) 308-0196				
rm PCT/IS/	A/210 (second sheet)(July 1992)±					

International application No. PCT/US96/01938

	i i		
C (Continua	tion). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant	passages	Relevant to claim No.
Y	Abstracts of the 1994 IDSA Annual Meeting, Clin. Infect. Volume 19, issued October 1994, R. Frothingham et al, "Sequence based strain differentiation in the Mycobacteriu tuberculosis complex, including rapid identification of M. BCG", page 565, see abstract 10.	ım	1-25
X	R. GHERNA et al, "AMERICAN TYPE CULTURE COLLECTION: CATALOGUE OF BACTERIA AND Pleighteenth edition, published 1992, pages 202 and 211, so document.	HAGES", ee entire	11-15
X	Infection and Immunity, Volume 62, No. 4, issued April	1994, L.	1-7, 16-25
Υ	Pascopella et al, "Use of in vivo complementation in Mycobacterium tuberculosis to identify a genomic fragme associated with virulence", pages 1313-1319, see entire d	nt ocument.	26
Y	Science, Volume 261, issued 10 September 1993, S. Arra "Cloning of an M. Tuberculosis DNA fragment associated entry and survival inside cells", pages 1454-1457, see end document.	d with	1-23
X	US,A,5,171,839 (PATARROYO) 15 December 1992, co	lumns 5-	1-10
Y	10.		16-23
Y	Nature, Volume 256, issued 07 August 1975, C. Kohler "Continuous cultures of fused cells secreting antibody of predefined specificity", pages 495-497, see entire documents		10
Y	US,A, 4,683,202 (MULLIS) 28 July 1987, see entire do	cument.	16-22, 24, 25
Y	Genomics, Volume 4, issued 1989, D. Wu et al, "The liamplification reaction (LAR) amplification of specific DI sequences using sequential rounds of template directed lipages 560-569, see figure 2.	NA	16-22, 24, 25
Y	US,A, 4,410,660 (STRAUS) 18 October 1983, columns 15.	14 and	23
Y	Gene, Volume 131, issued 1993, A. Kinger et al, "Identiand cloning of genes differentially expressed in the virul of mycobacterium tuberculosis", pages 113-117, see page column 2.	lent strain	1-26
X,P	WO,A2,95/17511 (JACOBS ET AL) 29 June 1995, see	entire	1-26

International application No. PCT/US96/01938

C (Continua	tion). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relev	ant passages	Relevant to claim No
X,E	J. Bacteriol., Volume 178, No. 5, issued March 1996, Mahairas et al, "Molecular analysis of genetic different mycobacterium bovis BCG and virulent M. bovis", pag 1282, see entire document.	ces between	1-26
Y, P	Microbiology, Volume 141, issued 1995, J. Rodriguez "Species-specific identification of mycobacterium bovis pages 2131-2138, see entire document.		1-7, 16-22, 24, 25
x	Hybridoma, Volume 13, No. 1, issued 1994, A. Arya	et al,	8-10
Υ	"Production and characterization of new murine monocantibodies reactive to mycobacterium tuberculosis", pag see page 27, table 1.		16-23
			Ç.
j			
			,
		:	

International application No. PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

C12Q 1/68; G01N 33/53; C12P 19/34; C12N 5/10, 1/21; C07K 5/00, 14/00, 16/00; C07H 21/02, 21/04

A. CLASSIFICATION OF SUBJECT MATTER: US CL :

435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.3, 24.32, 24.33

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE, BIOSIS, CAPLUS, WPIDS search terms: mycobacter?, tubercul?, bovis?, BCG, calmette, guerin, DNA, RNA, oligo, nucleic, oligonucleotide, hybridi?, probe, primer, amplif?, PCR, polymerase chain, ligase chain, LCR, attenuat?, immunoassay, antibod?, monoclon?, polyclon?, protein, peptide, antigen, virulenc?, infect?



CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12Q 1/68, G01N 33/53, C12P 19/34, C12N 5/10, 1/21, C07K 5/00, 14/00, 16/00, C07H 21/02, 21/04

A1

(11) International Publication Number:

WO 96/25519

(43) International Publication Date:

22 August 1996 (22.08.96)

(21) International Application Number:

PCT/US96/01938

(22) International Filing Date:

15 February 1996 (15.02.96)

(30) Priority Data:

08/390,878

17 February 1995 (17.02.95) US

(71) Applicant: PATHOGENESIS CORPORATION [US/US]; Suite 150, 201 Elliott Avenue West, Seattle, WA 98119 (US).

(72) Inventors: STOVER, Charles, Kendall; 7640 81st Place S.E., Mercer Island, WA 98040 (US). MAHAIRAS, Gregory, G.; 3312 39th West, Seattle, WA 98199 (US).

(74) Agents: HUNTER, Tom et al.; Townsend and Townsend and Crew, Steuart Street Tower, One Market, San Francisco, CA 94105-1492 (US). (81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

(54) Title: VIRULENCE-ATTENUATING GENETIC DELETIONS

(57) Abstract

The present invention provides specific genetic deletions that result in an avirulent phenotype of a mycobacterium. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
ВJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic		of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	Sī	Slovenia
CI	Côte d'Ivoire	Ll	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

10

15

20

25

30

VIRULENCE-ATTENUATING GENETIC DELETIONS

BACKGROUND OF THE INVENTION

Mycobacterium nuberculosis (MTB) infects over ten million people each year and kills over three million, making it the infectious agent causing the greatest mortality worldwide. In an effort to combat Mycobacterium nuberculosis, vaccination programs using a viable attenuated strain of Mycobacterium bovis called bacille Calmette-Guérin (BCG) have been established in more than 120 countries over the course of the last 5 decades. Although widely used and considered safe enough to administer to infants, the BCG vaccine is controversial for two principle reasons: 1) Efficacy for BCG vaccines against tuberculosis has varied from 0-85% in different clinical trials; and 2) Immunization with BCG sensitizes vaccinees to the tubercular antigens used in the tuberculin skin test, confounding attempts to discriminate between BCG immunization and TB infection. For these two reasons, especially the latter, BCG is not used in the United States where surveillance with the tuberculin test is preferred.

The original Pasteur BCG strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations which cannot revert. However, the mutations which arose during serial passage of the original BCG strain have never been identified. Moreover, recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli have also been unsuccessful again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence. The antigenicity of BCG and the characteristics leading to its avirulence are thus poorly understood.

SUMMARY OF THE INVENTION

The present invention provides specific genetic deletions that account for the avirulent phenotype of the bacille Calmette-Guérin (BCG) strain of *Mycobacterium bovis*. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

10

15

20

25

In a preferred embodiment, this invention provides for nucleic acid sequences that are markers for avirulent or virulent mycobacteria. The sequences uniquely characterize the presence or absence of deletions that result in an avirulent phenotype. More specifically the sequence are either deletion junction sequence or deletion sequences or subsequences within deletion junction sequences or deletion sequences. Thus, this invention provides for a marker for an avirulent mycobacterium comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement includes BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. In a particularly preferred embodiment, the marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, the marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The marker may be the full length BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3 or a subsequence within any of these regions. The marker may also include a nucleic acid having at least 80%, preferably 90%, more preferably 95%, and most preferably 98% percent sequence identity with BCGala, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3. The marker may also include a sequence selected from an open reading frame of a the deletion sequences BCGa1, BCGa2, BCGa3. Suitable open reading frames are indicated in Figures 4, 5, and 6.

The above described marker may be a probe. The probe may be labeled by a number of means including, but not limited to radioactive, fluorescent, enzymatic, and colorimetric labels.

In another embodiment, this invention provides for polypeptides encoded by a subsequence of the BCGa1, BCGa2, or BCGa3 deletions. In particular, the subsequence may be selected from an open reading frame (ORF) present in one of these deletion sequences. This invention also provides for monoclonal or polyclonal antibodies that

10

15

20

25

***** 30

specifically bind polypeptides encoded by one or more subsequences of the BCGa1, BCGa2, or BCGa3 deletions.

In still another embodiment, this invention provides for a recombinant cell comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGala, BCGalb, BCGala, BCGala

In still yet another embodiment, this invention provides a method of distinguishing between an attenuated and a virulent mycobacterium. The method involves detecting the presence or absence of a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGAla, BCGAlb, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGAlab, BCGA2ab, BCGA3ab, BCGA1b, BCGA2a, or BCGA3. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be

10

15

20

25

30

a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The polypeptide may be visualized by a number of means well known to those of skill in the art including antibody hybridization such as direct or indirect binding of labeled antibody.

This invention additionally provides a method for determining whether an attenuated or a virulent Mycobacterium is present in a sample. This method involves providing a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3; and hybridizing the first nucleic acid to the biological sample. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The method may also include detecting the hybridized first nucleic acid. This may involve direct detection of a label or additionally involve an amplification step and subsequent detection of the amplified product.

Finally, this invention provides a method of producing an attenuated-virulence mycobacterium. This method involves deleting from the genomic DNA of a virulent mycobacterium a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGA1, BCGA2, and BCGA3. The first nucleic acid may be BCGA1, BCGA2, or BCGA3, or alternatively, it may be a promoter, other control element or an open reading frame from BCGA1, BCGA2, or BCGA3.

Definitions

5

10

15

20

25

30

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The phrase "specifically detect" as used herein refers to the process of determining that a particular subsequence is present in a DNA sample. A DNA sequence may be specifically detected through a number of means known to those of skill in the art. These would include, but are not limited to amplification of the particular target sequence through polymerase chain reaction or ligase chain reaction, hybridization of the sequence to a labeled probe, and binding by labelled ligands or monoclonal antibodies. For a discussion of various means of detection of specific nucleic acid sequences see Perbal, B. A Practical Guide to Molecular Cloning, 2nd Ed. John Wiley & Sons, N.Y. (1988) which is incorporated herein by reference.

The phrase "select subsequence" is used herein to refer to a particular DNA subsequence that is of interest. It is often a predetermined or known sequence of nucleic acid bases. A select subsequence is typically chosen because of a unique sequence identity. Typically a select subsequence is targeted for DNA amplification and often is useful as a specific marker for the presence of a particular gene or a deletion of a particular nucleic acid sequence.

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides. Oligonucleotides may include, but are not limited to, primers, probes, nucleic acid fragments to be detected, and nucleic acid controls. Oligonucleotides include naturally occurring nucleotides, chemically modified naturally occurring nucleotides and synthetic nucleotides. The exact size of an oligonucleotide depends on many factors and the ultimate function or use of the oligonucleotide.

The term "primer" refers to an oligonucleotide, whether natural or synthetic, capable of acting as a point of initiation of DNA synthesis under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization (i.e., DNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. A primer is preferably a single-stranded oligodeoxyribonucleotide.

10

15

20

25

30

The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 25 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template.

The phrase "PCR primers competent to amplify" as used herein refers to a pair of PCR primers whose sequences are complementary to DNA subsequences immediately flanking the DNA subsequence (target sequence) which it is desired to amplify. The primers are chosen to bind specifically those particular flanking subsequences and no other sequences present in the sample. The PCR primers are thus preferably chosen to amplify the unique target sequence and no other. Alternatively, the PCR primers may be selected to bind to sequences other than the target sequence where the amplification products can be subsequently distinguished (e.g. where the desired amplified sequence is different in size than other amplified sequences).

"Amplifying" or "amplification", which typically refer to an "exponential" increase in target nucleic acid, are used herein to describe both linear and exponential increases in the number of a select target sequence of nucleic acid.

The term "antisense orientation" refers to the orientation of nucleic acid sequence from a structural gene that is inserted in an expression cassette in an inverted manner with respect to its naturally occurring orientation. When the sequence is double stranded, the strand that is the template strand in the naturally occurring orientation becomes the coding strand, and vice versa.

The term "deletion" refers to a region of a nucleic acid which is not present in an organism, but which is present in another related organism. In the context of mycobacteria, a deletion refers, e.g., to a region of nucleic acid which is not present in one strain of mycobacteria, but which is present in another related strain. For instance, an avirulent mycobacterial strain can have a deletion in its genome relative to the genome of a related virulent mycobacterial strain.

The term "deletion junction" refers to the region of a nucleic acid spanning the insertion point of a deletion. Thus, where a region of a nucleic acid sequence is deleted (i.e. a deletion is present), the deletion junction spans the nucleotides that are immediately adjacent to the deletion. Conversely, where a region of a nucleic acid sequence is not

deleted (i.e. the deletion is absent), two deletion junctions are present, each spanning respectively one end of the deletion sequence and its flanking sequence.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Figures 1, 2, or 3, or may comprise a complete cDNA or gene sequence.

10

15

5

Generally, a reference sequence is at least 10 nucleotides in length, frequently at least 20 to 25 nucleotides in length, and often at least 50 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

20

25

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch J. Mol. Biol. 48: 443 (1970); by the search for similarity method of Pearson and Lipman Proc. Natl. Acad. Sci. (USA) 85: 2444 (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected.

30

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned

10

15

20

25

30

sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The terms "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. The isolated nucleic acid probes of this invention do not contain materials normally associated with their *in situ* environment, in particular nuclear, cytosolic or membrane associated proteins or nucleic acids other than those nucleic acids intended to comprise the nucleic acid probe itself.

The term "marker" refers to a characteristic which distinguishes one class of cells or compositions from a second class of cells or compositions. For instance, the deletions and deletion junctions described herein can be used to distinguish between strains (e.g., virulent and avirulent strains) of mycobacteria. While markers are indicators of associated features or properties, as used herein, markers may also be used for purposes other than indicating the associated feature or property. Thus, for example, a nucleic acid marker of virulence identifies a particular nucleic acid which may be used in a variety of contexts other than simply indicating virulence.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "peptide" or "polypeptide" refers to an amino acid polymer which is encoded by a nucleic acid. The peptide or polypeptide may include naturally occurring or modified amino acids.

The terms "probe" or "nucleic acid probe" refer to a molecule that binds to a specific sequence or subsequence of a nucleic acid. A probe is preferably a nucleic acid which binds through complementary base pairing to the full sequence or to a subsequence of a target nucleic acid. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarily with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labelled such with, e.g., biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

5

10

15

20

25

30

The term "labeled nucleic acid probe" refers to a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen "bonds" to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference to a cell indicates that the cell replicates or expresses a nucleic acid, or expresses a peptide or protein encoded by DNA whose origin is exogenous to the cell. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes found in the native form of the cell wherein the genes are re-introduced into the cell by artificial means.

The term "sample" refers to a material with which bacteria may be associated. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. It will be recognized that the term "sample" also includes supernatant from eukaryotic cell cultures (which may contain free bacteria), cells from cell or tissue culture, and other media in which it may be desirable to detect mycobacteria (e.g., food and water).

The term "subsequence" in the context of a particular nucleic acid sequence refers to a region of the nucleic acid equal to or smaller than the specified nucleic acid.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least 90% sequence identity to a reference sequence, or preferably 95%, or more preferably 98% sequence identity to the

10

15

20

25

30

reference sequence, over a comparison window of at least about 10 to about 100 nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptides which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

The term "uninterrupted reading frame" or "open reading frame" refers to a DNA sequence (e.g., cDNA) lacking a stop codon or other intervening, untranslated sequence. An intact open reading frame refers to a full length uninterrupted reading frame or minor variations thereof.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal at a rate that is detrimental to the cell or animal within its host range. More particularly virulent mycobacteria persist longer in a host than avirulent mycobacteria. Virulent mycobacteria are typically disease producing and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism. Typically, infection of guinea pigs is used as an assay for mycobacterial virulence. In contrast, the term "avirulent" refers to a bacterium or strain of bacteria that either does not replicate within a host cell or animal within its host range, or replicates at a rate that is not significantly detrimental to the cell or animal.

The term BCG-like avirulence, as used herein refers to an attenuated virulence brought about by one of the deletions of the present invention.

10

15

20

25

30

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete sequence listing of the BCG deletion region 1 including flanking sequences. The deletion, designated BCG \(\Delta 1 \), is located between nucleotide 2327 and nucleotide 11126.

Figure 2 shows the complete sequence listing of the BCG deletion region 2 including flanking sequences. The deletion, designated BCG \(\Delta 2 \), is located between nucleotide 3382 and nucleotide 14071.

Figure 3 shows the complete sequence listing of the BCG deletion region 3 including flanking sequences. The deletion, designated BCG \(\Delta \)3, is located between nucleotide 1406 and nucleotide 10673. "N" represents "A", "C", "G", or "T".

Figure 4 shows a map of the deletion sequence BCGa1. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown.

Figure 5 shows a map of the deletion sequence BCGa2. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozomal binding sites and homologies to the predicted encoded proteins are shown.

Figure 6 shows a map of the deletion sequence BCGa3. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown. The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, remains to be determined. Therefore, the base pair coordinates given in the region 3 map 3' to the 9kb marker are approximations. The precise sequence determination of this region is likely to effect the length of open reading frames 3H and 3L.

Figure 7 illustrates the deletion junction regions of BCGa1, BCGa2, and BCGa3. The "terminal" deletion junction regions formed by the flanking sequences and the terminal regions of the deletion sequences are identified as BCGa1a, BCGa1b, BCGa2a, BCGa2b, and BCGa3a, and BCGa3b. When the deletion is present (the deletion sequences

10

15

20

25

30

are missing) the respective "a" and "b" sequences will be juxtaposed, thereby forming deletion "spanning" junction sequences designated BCG \triangle 1ab, BCG \triangle 2ab, and BCG \triangle 3ab, respectively.

Figure 8 shows EcoRI and BamHI restricted chromosomal DNAs from Mycobacterium bovis, BCG Connaught, and Mycobacterium tuberculosis strains H37Ra, H37Rv, and Erdman probed with ³²P labeled BCG subtracted probe.

DETAILED DESCRIPTION

This invention reflects the discovery of genetic deletions in mycobacteria that result in an avirulent genotype such as is exhibited by the bacille Calmette-Guérin (BCG) mycobacterium. The original Pasteur bacille Calmette-Guérin (BCG) strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations that cannot revert. The mutations that arose during serial passage of the original BCG strain were not previously known. Recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli were unsuccessful, again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence.

The genetic deletions leading to the avirulent phenotype of BCG were identified by genomic subtractions between Connaught strain of BCG and MBV/MTB. The subtracted probe resulting from the genomic subtraction between BCG and the H37 Rv strain of M. tuberculosis was subsequently used to identify and clone three regions from a cosmid library of Mycobacterium bovis genomic DNA. Southern blot mapping and DNA sequence comparisons between BCG and M. bovis showed that three regions, designated regions 1-3, contained DNA segments of approximately 9 kb, 11 kb and 9 kb respectively, which are deleted in the Connaught strain of BCG. Precise deletion junctions were identified for each region by comparisons of BCG and corresponding virulent MBV sequences. The respective deletions, designated BCGa1, BCGa2 and BCGa3 are illustrated in Figures 1-3.

One of skill in the art will appreciate that the deletions encompassed by BCGA1, BCGA2 and BCGA3 may be utilized in a variety of contexts. For example, the deletions may be utilized to distinguish between avirulent and virulent strains of

10

15

20

25

30

mycobacteria thereby providing early detection of patients at risk for tuberculosis. This is of particular importance where mycobacteria are identified in a sample from a patient that has been previously vaccinated with BCG. In this context it may be critical to determine whether mycobacteria identified in a biological sample from such a patient are pathogenic.

13

In another embodiment, the preparation of mycobacteria containing the deletions of the present invention may provide superior vaccines to BCG which has long been known to have marginal efficacy. Thus, for example, a *Mycobacterium tuberculosis* may contain a full BCGal deletion or a smaller deletion within BCGal (e.g. one or more open reading frames) rendering it avirulent. An avirulent MTB will provide a more efficient vaccine because it is antigenically more similar to MTB than is BCG. Moreover, an MTB rendered avirulent by the production of smaller deletions within the deletion regions identified in this invention will present more antigenic determinants.

Since the loss of virulence is due to the loss of gene products expressed by the nucleic acid sequences comprising the deletion regions, the BCGA1, BCGA2 and BCGA3 deletion sequences and proteins encoded within these deletion sequences provide suitable targets for drug screening. Thus, the use of deleted sequences as targets to screen for drugs that inhibit or interfere with transcription, translation, or post-translational processing of proteins encoded by the deletion sequences, or with the deletion encoded polypeptides themselves, provides an assay for anti-mycobacterial agents. In particular, the use of reporter genes such as firefly luciferase (FFlux), \(\beta\)-galactosidase (BGal), and the like, under the control of promoters present in the deletion sequence provide a rapid assay for drugs regulating activity originating in this region. Conversely, since the protein products of the deletion sequences are presumably expressed in virulent mycobacterial species, proteins expressed by deletion sequences may make good antigens for antimycobacterial vaccines.

Finally, as the viability of BCG demonstrates, deletion regions BCGA1, BCGA2 and BCGA3 are not required for mycobacterial growth and reproduction. Thus, these deletion regions provide good insertion points for the expression of heterologous DNA. The heterologous DNA sequences may be under the control of endogenous inducible or constitutive promoters typically found in the deletion sequences, or alternatively, they may be under the control of introduced promoters, either constitutive or inducible, exogenous to mycobacteria.

10

15

20

I. Detection of Deletions

As indicated above, the deletions identified in the present invention provide useful markers for the identification of an avirulent (or conversely a virulent) mycobacterial phenotype. Specifically, determination of avirulence simply requires the detection of the presence or absence of the deletion (either BCGa1, BCGa2, or BCGa3, or deletions within these regions). Where the deletion is present in the bacterial DNA, the bacterium expresses a BCG-like avirulent phenotype. Conversely, where the deletion is absent in the bacterial DNA, the bacterium does not express a BCG-like avirulence. While this may indicate that the bacterium is virulent, one of skill will appreciate that the bacterium may still be avirulent due to the presence of other mutations or deletions. Nevertheless, screening for the presence of the deletion provides a means of detecting a BCG-like avirulent mycobacterium.

Means of detecting deletions are well known to those of skill in the art. Generally, the deletions may be detected either by detecting the presence or absence of deletion junctions, or, alternatively, by detecting the presence or absence of the sequences contained within the deletion (deletion sequences). Where a nucleic acid sequence is deleted (i.e., a deletion is present), the sequences that previously flanked the deleted sequence are juxtaposed, thereby forming a new deletion junction that spans the deletion. Detection of the presence of such a "spanning" deletion junction indicates the presence of the deletion and thus the avirulent phenotype.

Conversely, where the nucleic acid sequence is not deleted (the deletion is not present) the spanning junction sequence will be absent (See, e.g. Figure 7). The "terminal" deletion junction sequences flanking each endpoint of the deletion region are present and detection of these terminal deletion junctions indicates the absence of a deletion. Spanning deletion junction regions and terminal deletion junctions suitable for detecting the deletions of the present invention are illustrated in Figure 7 and in Table 1.

Table 1. Nucleic acid sequences comprising deletion junctions. The symbol "|" indicates the insertion point of the deletion sequence. Deletion sequence bases are represented in lower case letters.

Junction	Nucleotide Sequence	Seq. ID
BCG∆la	CTGGTCGACGATTGGCACAT gcagccgtgggtgccgccgg	1

25

T		
BCG∆1b	gtgtcttcatcggcttccac CCAGCCGCCCGGATCCAGCA	2
BCG∆2a	CAACTCCACGGCGACCACCC gcgcccccgctcgcactaga	3
BCG∆2b	gcccacccggtcgagcaccc CGATGATCTTCTGTTTGACC	4
BCG∆3a	CACCTCGACCACGGCCAACC gtggacctgtgagatacact	5
BCG∆3b	tcagcagtccacggccaacc CCGCACCAACACCTTCCACC	6
BCG∆1ab	CTGGTCGACGATTGGCACAT CCAGCCGCCCGGATCCAGCA	7
BCG∆2ab	CAACTCCACGGCGACCACCC CGATGATCTTCTGTTTGACC	8
BCG∆3ab	CACCTCGACCACGGCCAACC CCGCACCAACACCTTCCACC	9

15

5

Where a deletion is detected by determining the presence or absence of sequences contained within the deletion (deletion sequences), the absence of deletion sequences indicates the presence of a deletion and thus an avirulent phenotype. Conversely, the presence of deletion sequences indicates the absence of a deletion. Deletion sequences that provide suitable targets for detecting the deletions of the present invention are provided in Figures 1, 2 and 3.

A) Isolation of DNA for Detection of Mycobacterium Genomic Deletions

20

In a preferred embodiment, DNA is obtained from mycobacteria. As used herein, the term "mycobacteria" refers to any bacteria of the family Mycobacteriaceae (order Actinomycetales) and includes, but is not limited to, Mycobacterium tuberculosis, Mycobacterium avium complex, Mycobacterium kansasii, Mycobacterium scrofulaceum, Mycobacterium bovis and Mycobacterium leprae. These species and groups and others are described in Baron, S., ed. Medical Microbiology, 3rd Ed. (1991) Churchill Livingstone, New York, which is incorporated herein by reference.

25

The identification of deletions using a DNA marker requires that the DNA sequence be accessible to the particular probes used or to the components of the amplification system if the DNA sequence is to be amplified. In general, this accessibility is ensured by isolating the nucleic acids from the sample.

30

A variety of techniques for extracting nucleic acids from biological samples are known in the art. For example, see those described by Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New

10

15

20

25

30

York, (1985), by Han, et al. Biochemistry, 26: 1617-1625 (1987) and by Du, et al. Bio/Technology, 10: 176-181 (1992), which are incorporated herein by reference.

Alternatively, if the sample is readily disruptable, the nucleic acid need not be purified prior to amplification by the PCR technique, *i.e.*, if the sample is comprised of cells, particularly peripheral blood lymphocytes or monocytes, lysis and dispersion of the intracellular components may be accomplished merely by suspending the cells in hypotonic buffer or boiling them in a low concentration of alkali (*i.e.* 10 mM NaOH).

In a preferred embodiment, DNA is extracted from mycobacteria as described in Example 1.

B) Detection of Deletions Using Hybridization Probes

In one embodiment the avirulence deletions are detected by contacting DNA obtained from the mycobacterium with a probe that specifically binds an entire deletion junction region or a subsequence of that region and does not specifically bind to any other DNA sequences in the sample. Alternatively, a probe that specifically binds the entire deleted region or subsequence of that region and does not specifically bind to any other sequences in the sample is also suitable. While such probes may be proteins, oligonucleotide probes are preferred. Typically, the sequence of the oligonucleotide probe is chosen to be complementary to a select subsequence unique to the deletion junction or the deletion sequence, whose presence or absence is to be detected. Under stringent conditions the probe will hybridize with the select subsequence forming a stable duplex.

The probe is typically labeled. Detection of the label in association with the target DNA indicates either the presence or absence of the deletion. The probe may be used to detect the deletion junction or deletion sequences directly in a DNA sample without amplification of the deletion subsequences. In one embodiment, unamplified DNA sequences are probed using a Southern blot. The DNA of the sample is immobilized, on a solid substrate, typically a nitrocellulose filter or a nylon membrane. The substrate-bound DNA is then hybridized with the labeled probe under stringent conditions and non-specifically hybridized probe is washed away. Labeled probe detected in association with the immobilized mycobacterial sequences (e.g. bound to the substrate) indicates the presence of deletion sequences (e.g. BCG\Delta1, BCG\Delta2, or BCG\Delta3) and therefore the absence of the deletion. Means for detecting specific DNA sequences are well known to those of skill in

the art. Protocols for Southern blots as well as other detection methods are provided in Maniatis, et al. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY (1982), which is incorporated herein by reference.

In another embodiment, the mycobacterial DNA subsequences are themselves labeled. They are then hybridized, under stringent conditions, with a probe immobilized on a solid substrate. Detection of the label in association with the immobilized probe indicates the presence or absence of the deletion.

In a preferred embodiment, the deletion junction sequences or subsequences or the deletion sequences or subsequences may be amplified by a variety of DNA amplification techniques (for example via cloning, polymerase chain reaction, ligase chain reaction, transcription amplification, etc.) prior to detection using a probe. Because the copy number of mycobacterial sequences bearing the virulence-attenuating deletions is low, the use of unamplified mycobacterial DNA results in an assay of low sensitivity. Amplification of mycobacterial DNA increases sensitivity of the assay by providing more copies of possible target subsequences. In addition, by using labeled primers in the amplification process, the mycobacterial DNA sequences are labeled as they are amplified.

C) Selection of Probes for Detection of the Deletion Junction Sequences or the Deletion Sequences

20

25

5

10

15

Full length sequences are provided for the deletions BCGa1, BCGa2, and BCGa3 in Figures 1, 2 and 3 respectively. Using these sequence listings, one of skill in the art may easily determine appropriate probes or primers for the detection of the presence or absence of the deletion junctions or the deletion sequences. Generally speaking, a probe will be selected that hybridizes to the target junction sequences or deletion sequences, but not to other mycobacterial nucleic acid sequences under stringent conditions. The design of hybridization probes is well known in the art. See, for example, Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989), which is incorporated herein by reference.

30

In a preferred embodiment, the probe is an oligonucleotide sequence complementary to a subsequence comprising a deletion junction (e.g. BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b, BCGΔ1ab, BCGΔ2ab, and BCGΔ3ab) or a

10

15

20

25

30

sequence complementary to a subsequence of a deletion sequence (e.g. BCG Δ 1, BCG Δ 2, and BCG Δ 3). The probe preferably has destabilizing mismatches with subsequences from other regions of the mycobacterial genome.

The exact length of the probe depends on many factors including the length of conserved regions around the deletions, the degree of sequence specificity desired, and the amount of internal complementarity within the probe. Such probes are preferably 17 to 25 bases in length. One of skill will recognize that longer probes specifically hybridize at higher temperatures. Generally, stringent conditions are selected to be about 5°C to 20°C, more preferably about 10°C, lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from an avirulent mycobacterium such as BCG, but not to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV. Alternatively, Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from an avirulent mycobacterium such as BCG.

Oligonucleotide probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang et al. Meth. Enzymol, 68: 90-99 (1979); the phosphodiester method of Brown et al., Meth. Enzymol. 68:109-151 (1979); the diethylphosphoramidite method of Beaucage et al., Tetra. Lett., 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

Probe detectability may be increased by the attachment of a label. As used herein, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g. DynabeadsTM), fluorescent dyes (e.g., fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads.

Methods for attaching labels to probes, primers, and antibodies are well known to those of skill in the art. For example, the probe can be labeled at the 5'-end with ³²P by incubating the probe with ³²P-ATP and polynucleotide kinase (see Perbal, A

Practical Guide to Molecular Cloning, 2nd ed. John Wiley, N.Y. (1988)). Other labels may be joined to the probe directly or through linkers. They may be located at the ends of the probe or internally. Methods of attaching labels may be found in Connell, et al., Bio/Techniques 5: 342 (1987), U.S. Patent Nos. 4,914,210, 4,391,904 and 4,962,029, which are incorporated herein by reference. In addition, kits for labelling oligonucleotides are widely available. See, for example, Boehringer Mannheim Biochemicals (Indianapolis, IN) for "Genius" labeling kits based on dioxigenin technology and Clonetech (South San Francisco, CA) for a variety of direct and indirect oligonucleotide labeling reagents.

10

15

20

25

30

5

D) Detection of Deletions Conferring Avirulence Through Amplification of Unique Subsequences

Deletions are particularly amenable to detection without the use of a hybridization probe. In a preferred embodiment, subsequences are amplified that include a deletion junction. The amplified deletion junction may be a "spanning" deletion junction in which case where the deletion is present (i.e. the deletion sequences are absent), the amplification product is a specific DNA incorporating the deletion junction sequence spanning the deletion (e.g. incorporating flanking sequences from both sides of the deleted sequence). Where the deletion is absent (i.e. deletion sequences are present) and primers are selected so that there are no priming sites within the deletion sequences, amplification is non-existent or alternatively provides a complex mixture of non-specifically amplified fragments. Alternatively, amplification primers may be selected that specifically hybridize to deletion sequences, as long as they are selected to amplify sequences that are distinguishable from the sequence amplified when the deletion is present.

Alternatively, the amplification product may be subsequence of a "terminal" deletion junction in which case absence of the deletion (i.e. the deletion sequences are present) will result in the amplification of the specifically targeted nucleic acid. Conversely, where the deletions are present (i.e. the deletion sequences are absent) there will be no specific amplification of a terminal deletion junction.

Amplification products may be separated by size for characterization. Size separation may be accomplished by a variety of means known to those of skill in the art.

10

15

20

25

30

These methods include, but are not limited to electrophoresis, density gradient centrifugation, liquid chromatography, and capillary electrophoresis. In a preferred embodiment, the fragments are separated by agarose gel electrophoresis. The bands are then stained with a marker to visualize them such as ethidium bromide and the gel is visualized, e.g., using ultraviolet light.

As described above, an agarose gel typically shows 1 band if the deletion is present, reflecting amplification of the deletion-spanning sequence. Where the deletion is absent, amplification results in either no bands, where there are no sequences within the deletion to which the amplification primers may hybridize, or a smear where there is non-specific amplification, or a series of discrete bands distinguishable from the band representing the deletion-spanning sequence where primers are chosen that hybridize to deletion sequences.

E) Selection of Primers for Amplification of Avirulence Deletions

Amplification of deletion junction sequences or subsequences or deletion sequences or subsequences may be accomplished by methods well known in the art, which include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990), which is incorporated herein by reference), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), which are incorporated herein by reference), transcription amplification (see Kwoh, et al., Proc. Natl. Acad. Sci. (U.S.A.), 86: 1173 (1989) which is incorporated herein by reference), and self-sustained sequence replication (see Guatelli, et al., Proc. Nat. Acad. Sci. (U.S.A.), 87: 1874 (1990) which is incorporated herein by reference), each of which provides sufficient amplification so that the target sequence can be detected by nucleic acid hybridization to a probe or by electrophoretic separation. Alternatively, methods that amplify the hybridization probe to detectable levels can be used, such as $Q\beta$ -replicase amplification. See, for example, Kramer, et al. Nature, 339: 401 (1989), Lizardi, et al. Bio/Technology, 6: 1197 (1988), and Lomell, et al., Clin. Chem. 35: 1826 (1989) which are incorporated herein by reference.

In a preferred embodiment, amplification is by polymerase chain reaction using a pair of primers that flank and thereby amplify a selected deletion junction subsequence. Selection of primers is readily apparent to one of skill in the art using the sequence listings of the present invention. For example, a pair of PCR primers 5'-TCGACGATTGGCACAT-3' ($T_m = 55^{\circ}$ C) and 5'-TCCCTCCCTGTATTTGTAT-3' ($T_m = 56^{\circ}$ C) will amplify a 469 base pair sequence including the BCGala deletion junction, while 5'-CGTTCTTCGGAGGTTTC-3' ($T_m = 56^{\circ}$ C) and 5'-GGCGGCTGGGTGGA-3' ($T_m = 60^{\circ}$ C) will amplify a 471 base pair sequence including the BCGalb deletion junction.

10

15

20

25

30

5

F) Detection of Deletions through Detection of Expression Products of Deletion Sequences

In addition to the detection of deletions by the detection of either the deletion junction sequences or the deletion sequences, one may detect the absence of the deletion by detecting the expression products of the deletion sequences. Thus, for example, where the deletion sequences express a protein, the presence of that protein indicates the absence of the deletion and thus is indicative of a virulent (non BCG-like) phenotype. Such proteins are referred to herein as "deletion polypeptides".

Means of determining proteins expressed by particular nucleic acid sequences are well known to those of skill in the art. Typically this involves determining the longest open reading frame. This may be aided by the identification of initiation sites (e.g. ribozome binding sites). The protein encoded by the largest open reading frame is determined using codon preferences for the specific organism from which the nucleic acid is obtained. The polypeptide sequence listing may then be compared against a sequence database, e.g. GenBank, to determine other sequences sharing substantial sequence identity with the calculated sequence. The expression of the protein may be verified by isolating and then sequencing proteins having the predicted length and charge characteristics.

Once deletion polypeptides are identified they may be detected by routine methods well known to those of skill in the art. Typically this involves isolating and then detecting the polypeptide. The polypeptide may be isolated by a number of means well known to those of skill in the art. This includes typical methods of protein

10

15

20

25

30

purification such as high performance liquid chromatography (HPLC), electrophoresis, capillary electrophoresis, hyperdiffusion chromatography, thin layer chromatography, and the like. Methods of purifying and detecting proteins are well known to those of skill in the art (see, e.g., Methods in Enzymology Vol. 182: Guide to Protein Purification, M. Deutscher, ed. Vol. 182 (1990), which is incorporated herein by reference).

Alternatively, deletion polypeptides sequences may be detected using immunoassays utilizing antibodies specific for the deletion polypeptides. The production of such antibodies and their use in immunoassays is detailed below.

G) Antibodies to Deletion Polypeptides

Antibodies can be raised to the polypeptides encoded by the nucleic acids corresponding to the open reading frames present in the deletion regions of the present invention (deletion polypeptides). As used herein "antibodies" include immunoglobulin or a population of immunoglobins which specifically bind to an antigen. Thus an antibody may be monoclonal or polyclonal including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies can be raised to these polypeptides in either their native configurations or in non-native configurations. Anti-idiotypic antibodies may also be used.

1) Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with deletion polypeptides. Recombinant polypeptides are the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring polypeptides may also be used either in pure or impure form. Synthetic peptides made using sequences described herein may also used as immunogens for the production of antibodies.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide is injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

10

15

20

25

30

Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified deletion polypeptide is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY, which are incorporated herein by reference.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256: 495-497, which discusses one method of generating monoclonal antibodies.

Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells (See, Kohler and Milstein (1976) Eur. J. Immunol. 6: 511-519, incorporated herein by reference). The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro.

Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al. (1989) Science 246: 1275-1281. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B

10

15

20

25

30

cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, Huse et al. Science 246: 1275-1281 (1989); and Ward, et al. Nature 341: 544-546 (1989). The polypeptides and antibodies of the present invention are used with or without modification, including chimeric antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen et al. Proc. Nat'l Acad. Sci. USA 86: 10029-10033 (1989).

Antibodies, including binding fragments and single chain versions, against predetermined fragments of deletion polypeptides can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptides, or screened for agonistic or antagonistic activity, e.g., mediated through a receptor. These monoclonal antibodies will usually bind with at least a K_D of about 1 mM, more usually at least about 300 μ M, and most preferably at least about 0.1 μ M or better.

The antibodies of this invention can also be used for affinity chromatography in isolating deletion polypeptides. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a bacterial lysate, or recombinant cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified deletion polypeptides are released.

10

15

20

25

30

The antibodies can be used to screen expression libraries for particular expression products. Usually the antibodies in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

In a preferred embodiment, antibodies to deletion polypeptides are used for the identification of cell populations expressing the polypeptides. By assaying the expression products of cells expressing the polypeptides it is possible to diagnose bacterial infections.

Antibodies raised against each polypeptide are useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to the presence of the respective antigens.

2) Immunoassays

A particular deletion polypeptide can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) 1991 Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in Maggio (ed.) (1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is incorporated herein by reference. See also Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed.) (1988) Non-isotopic Immunoassays Plenum Press, NY.

Immunoassays for measurement of deletion polypeptides can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be, e.g., competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with a deletion polypeptide produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

10

15

20

25

30

In a competitive binding immunoassay, the deletion polypeptide present in the sample competes with labelled protein for binding to a specific binding agent, for example, an antibody specifically reactive with a particular deletion polypeptide. The binding agent is, e.g., bound to a solid surface to produce separation of bound labelled polypeptide from the unbound labelled polypeptide. Alternately, the competitive binding assay may be conducted in liquid phase and any of a variety of techniques known in the art may be used to separate the bound labelled protein from the unbound labelled protein. Following separation, the amount of bound labeled protein is determined. The amount of polypeptide present in the sample is inversely proportional to the amount of labelled polypeptide binding.

Alternatively, a homogenous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding agent. This alteration in the labelled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the polypeptide.

Deletion polypeptides may also be detected by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which is also an antibody, and which binds the protein at a different site, is labelled. After binding at both sites on the protein, the unbound labelled binding agent is removed and the labelled binding agent bound to the solid phase is measured. The amount of labelled binding agent bound is directly proportional to the amount of polypeptide in the sample.

Western blot analysis can be used to determine the presence of a deletion polypeptide in a sample. Electrophoresis is carried out, for example, on a bacterial sample suspected of containing the deletion polypeptide. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support such as a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody is labelled, or alternatively may be it is detected by subsequent incubation with a second labelled antibody that binds the primary antibody.

10

15

20

25

30

The immunoassay formats described above employ labelled assay components. The label can be in a variety of forms as described above. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with polypeptides include competitive and noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant deletion polypeptide as described above. Other sources of polypeptides, including isolated or partially purified naturally occurring protein, can also be used. Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labelled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can be also be used similar to those described above for the measurement of deletion polypeptides.

II. Preparation of Deletion-Containing Mycobacteria

Mycobacteria containing specific deletions may be prepared by using methods of homologous recombination well known to those of skill in the art. In brief, homologous recombination is a natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar (i.e. "homologous") sequences, and the ligation of the two molecules such that one region of each initially

10

15

20

25

30

present molecule is now ligated to a region of the other initially present molecule (Sedivy, *Bio/Technol.*, 6: 1192-1196 (1988).

Homologous recombination is exploited by a number of various methods of "gene targeting" well known to those of skill in the art. (see, for example, Mansour et al. Nature, 336: 348-352 (1988); Capecchi Trends Genet. 5: 70-76 (1989); Capecchi Science 244: 1288-1292 (1989); Capecchi et al. pages 45-52 In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman et al. Cell 56: 145-147 (1989)). Some approaches focus on increasing the frequency of recombination between two DNA molecules by treating the introduced DNA with agents which stimulate recombination (e.g. trimethylpsoralen, UV light, etc.), however, most approaches utilize various combinations of selectable markers to facilitate isolation of the transformed cells.

One such selection method is termed positive/negative selection (PNS) (Thomas and Cappechi Cell 51: 503-512 (1987)). This method involves the use of two selectable markers: one a positive selection marker such as the bacterial gene for neomycin resistance (neo'); the other a negative selection marker such as the herpes virus thymidine kinase (tk) gene. Neo' confers resistance to the drug G-418, while herpes tk renders cells sensitive to the nucleoside analog gangcyclovir (GANC) or 1-(2-deoxy-2-fluoro-b-d-arabinofuranosyl)-5-iodouracil (FIAU). The DNA encoding the positive selection marker in the transgene (e.g. neo^R) is generally linked to an expression regulation sequence that allows for its independent transcription in mycobacteria. It is flanked by first and second sequence portions of at least a part of the deletion or deletion flanking sequences.

These first and second sequence portions target the transgene to a specific nucleotide sequence. A second independent expression unit capable of producing the expression product for a negative selection marker, e.g. for herpes virus tk is positioned adjacent to or in close proximity to the distal end of the first or second portions of the first DNA sequence. Upon transfection, some of the mycobacteria incorporate the transgene by random integration, others by homologous recombination between the endogenous allele and sequences in the transgene. As a result, one copy of the targeted nucleic acid is disrupted by homologous recombination with the-transgene with simultaneous loss of the sequence encoding herpes tk gene. Random integrants, which

10

15

20

25

30

occur via the ends of the transgene, contain herpes tk and remain sensitive to GANC or FIAU. Therefore, selection, either sequentially or simultaneously with G418 and GANC enriches for transfected mycobacteria containing the transgene integrated into the genome by homologous recombination.

Methods of homologous recombination in mycobacteria are described in greater detail by Ganjam et al. Proc. Natl. Acad. Sci. USA, 88: 5433-5437 (1991) and Aldovini et al., J. Bacteriol., 175: 7282-7289 (1993) which are incorporated herein by reference.

III. Screening for Drug Susceptibility/Therapeutics

The expression products of the open reading frames in the BCGA1, BCGA2, and BCGA3 deletions of the present invention are targets for anti-mycobacterial drugs. To determine particularly suitable drug targets, open reading frames and surrounding expression control sequences are introduced into avirulent strains of mycobacteria, alone or in combination with other open reading frame regions to determine which regions are critical for virulence. Once particular genes are identified as critical for virulence, anti-mycobacterial agents are designed to inhibit expression of the critical genes, or to attack the critical gene products. For instance, antibodies are generated against the critical gene products and used as prophylactic or therapeutic agents. Alternatively, small molecules can be screened for the ability to selectively inhibit expression of the critical gene products, e.g., using recombinant expression systems which include the gene's endogenous promoter. These small molecules are then used as therapeutics, or prophylactic agents to inhibit mycobacterial virulence.

In another embodiment, anti-mycobacterial agents which render a virulent mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium.

Such transformed mycobacteria are useful as vaccine components, and as components of immunological infectivity assays. For instance, an animal's blood can be monitored for the presence of anti-mycobacterial antibodies using the procedures described herein, using transformed avirulent mycobacterial components in various

10

15

20

25

30

immunological assays. Anti-mycobacterial agents useful in this invention include, without limitation, antisense genes, ribozymes, decoy genes, transdominant proteins and suicide genes.

An antisense nucleic acid is a nucleic acid that, upon expression, hybridizes to a particular mRNA molecule, to a transcriptional promoter or to the sense strand of a gene. By hybridizing, the antisense nucleic acid interferes with the transcription of a complementary DNA, the translation of an mRNA, or the function of a catalytic RNA. Antisense molecules useful in this invention include those that hybridize to gene transcripts in the region of the deletions of the invention, particularly deletion region 1.

A ribozyme is a catalytic RNA molecule that cleaves other RNA molecules having particular nucleic acid sequences. Ribozymes useful in this invention are those that cleave deletion gene transcripts. Examples include hairpin and hammerhead ribozymes.

A decoy nucleic acid is a nucleic acid having a sequence recognized by a regulatory DNA binding protein (i.e., a transcription factor). Upon expression, the transcription factor binds to the decoy nucleic acid, rather than to its natural target in the genome. Useful decoy nucleic acid sequences include any sequence to which a transcription factor binds in the deletion regions of the present invention.

A transdominant protein is a protein whose phenotype, when supplied by transcomplementation, will overcome the effect of the native form of the protein. For instance, an avirulent mycobacterium can be rendered virulent by introducing transdominant proteins from deletion region 1.

A suicide gene produces a product which is cytotoxic. In the vectors of the present invention, a suicide gene is operably linked to an inducible expression control sequences which is stimulated upon infection of a cell by a mycobacterium.

IV. Use of Expressed "Deletion Proteins" in a Vaccine

The deletion polypeptides encoded by the open reading frames in BCGa1, BCGa2, and BCGa3 may be recombinantly expressed and used as components of immunological assays as described above or in vaccines. Expression of polypeptides

encoded by the open reading frames of the BCGa1, BCGa2, or BCGa3 deletions may be accomplished by means well known to those of skill in the art.

In brief, the expression of natural or synthetic nucleic acids encoding deletion polypeptides will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of polynucleotide sequence encoding deletion polypeptides.

10

15

5

To obtain high level expression of a cloned gene, such as those polynucleotide sequences encoding deletion polypeptides, it is desirable to construct expression plasmids which contain, at the minimum, a promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, *i.e.*, shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. For detailed techniques employed in the recombinant expression of deletion proteins see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual (2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory (1989)), Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques (Berger and Kimmel (eds.), San Diego: Academic Press, Inc. (1987)), or Current Protocols in Molecular Biology, (Ausubel, et al. (eds.), Greene Publishing and Wiley-Interscience, New York (1987), all of which are incorporated herein by reference.

25

30

20

The expressed deletion polypeptides may be used in a variety of assays. For example, the deletion polypeptides can be used as reagents in immunoblot assays to test whether a patient was previously exposed to virulent mycobacteria (i.e., to test whether the patient has antibodies to the deletion polypeptide). These assays have the advantage of discriminating between previous exposure to an avirulent mycobacterium (e.g., one used in a vaccine) and exposure to a virulent mycobacterium. Thus, vaccinated individuals can be tested for antibodies to the virulent mycobacterium without regard to whether the patient has been vaccinated with an avirulent mycobacterium.

10

15

20

25

30

The deletion polypeptides can also be used as antigenic vaccine components to direct antibodies to elements which are critical for virulence. These polypeptides can be added to existing vaccines (e.g., those based upon avirulent mycobacteria and which lack the deletion polypeptide) to supplement the range of antigenicity conferred by the vaccine, or they may be used apart from other mycobacterial antigens. The vaccines of the invention contain as an active ingredient an immunogenically effective amount of a deletion polypeptide or of a recombinant vector which includes the deletion polypeptide. The immune response can include the generation of antibodies; activation of cytotoxic T lymphocytes (CTL) against cells presenting peptides derived from the polypeptides or other mechanisms well known in the art. See e.g. Paul Fundamental Immunology Third Edition published by Raven press New York (incorporated herein by reference) for a description of immune response. Useful carriers are well known in the art, and include, for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, and polyamino acids such as poly(D-lysine:D-glutamic acid). The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The compositions are suitable for single administrations or a series of administrations. When given as a series, inoculations subsequent to the initial administration are given to boost the immune response and are typically referred to as booster inoculations.

The vaccine compositions of the invention are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration that comprise a solution of the agents described above dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile

٠. >

÷

Ş.

5

10

15

20

25

30

solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient and more preferably at a concentration of 25%-75%.

For aerosol administration, the polypeptides are preferably supplied in finely divided form along with a surfactant and propellant. The surfactant should be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

The amount of vaccine administered to the patient will vary depending upon the composition being administered, the physiological state of the patient and the manner of administration.

Live attenuated recombinant viruses which include the deletion polypeptide, such as recombinant vaccinia or adenovirus vectors, are convenient alternatives as vaccines because they are inexpensive to produce and are easily transported and administered. Vaccinia vectors and methods useful in immunization protocols are described, for example, in U.S. Patent No. 4,722,848, incorporated herein by reference.

Deletion sequences and subsequences of this invention may also be used in methods of genetic immunization. Briefly, genetic immunization involves transfecting

10

15

20

25

30

cells in vivo with nucleic acids encoding pathogen specific antigens. The transformed host cells then express the antigen thereby stimulating the host immune system.

In the present invention, antigen-encoding deletion region sequences are used to transform mammalian host cells thereby resulting in the expression of the antigen by the host. This provokes an immune response by the host against the expressed antigen thereby conferring immunity on the host. Methods of genetic immunization are well known to those of skill in the art (see, e.g., Wang et al. Proc. Natl. Acad. Sci. USA, 90: 4156-4160 (1993); Ulmer et al., Science, 259: 1745-1749 (1993); Fynan et al. DNA Cell Biol., 12: 785-789 (1993); Fynan et al. Proc. Natl. Acad. Sci. USA, 90: 11478-11482 (1993); Robinson et al. Vaccine, 11: 957-960 (1993); and Martinon et al. Eur. J. Immunol., 23: 1719-1722 (1993), which are incorporated herein by reference.

VI. Use of Promoters within Deletion Sequences for Expression of Recombinant Proteins

Bacille Calmette-Guérin (BCG) contains all three deletions (BCGΔ1, BCGΔ2, and BCGΔ3) and yet is able to grow and reproduce indicating that the sequences contained within the deletion are not essential for bacterial viability. These deletion regions therefore make good target sites for the insertion of heterologous DNA as mycobacteria are tolerant of disruption of the native genome in these regions. The BCGΔ1, BCGΔ2, and BCGΔ3 deletion regions therefore provide suitable target sites for the incorporation of expression cassettes and the subsequent expression of exogenous gene products. The expression cassettes typically comprise a nucleic acid sequence under the control of a promoter. The promoter may be either constitutive or inducible. The cassette may additionally comprise a selectable marker such as an antibiotic resistance gene, a gene encoding a fluorescent marker (e.g. green fluorescent protein), or a gene encoding an enzymatic marker (e.g. β-galactosidase).

Alternatively, genes under the control of endogenous promoters may be used as well. In one embodiment, reporter genes under the control of endogenous promoters found within the deletion sequences may be inserted at the deletion sites. These reporter genes may be utilized as an assay for antimycobacterial compounds that act by inhibiting transcription or translation of deletion sequences. Assaying for the

10

15

20

25

30

reporter gene product in the presence of an antimycobacterial compound provides a measure of efficacy of that compound in upregulating or downregulating deletion sequence genes. Methods of use of mycobacterial reporter gene assays to screen for drug activity are described by Cooksey et al., Antimicrob. Agents Chemother., 37: 1348-1352 (1993), and Jacobs et al., Science, 260: 819-822 (1993) which are incorporated herein by reference.

EXAMPLES

The following examples are offered by way of illustration, not by way of limitation.

Example 1

Identification of Virulence-Attenuating Deletions

Bacterial Culture

All strains of Mycobacteria used in this study were maintained in 7H9 (Difco, Detroit Michigan, USA) media supplemented with OADC (BBL) or were grown on 7H11 agar supplemented with oleic acid albumin dextrose complex (OADC). Escherichia coli (strain DH5 α or NM554) was used as a host for all recombinant plasmids and cosmids. E. coli was maintained in LB medium with or without agar. Carbenicillin (100 μ g/ml) was used in place of ampicillin for the selection of all E. coli plasmids.

Extraction of High Molecular Weight DNA

High molecular weight chromosomal DNA was prepared by diluting a late log phase culture of the respective mycobacterium 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing growth for 4 to 5 days. The cells were then harvested by centrifugation, washed once in TE (pH 8.0) and resuspended in 4 ml of 25% sucrose in 10X TE. 100 μ g of lysozyme was added and the preparation was incubated at 37°C for 2 hr followed by the addition of 100 μ g of proteinase K and sarkosyl to a concentration of 1% weight/volume. Following overnight incubation at 65°C the mixture was extracted 4 times with chloroform isoamyl alcohol 24:1, once with phenol/chloroform (1:1), and twice again with chloroform isoamyl alcohol. The resulting high molecular weight DNA was then run on a CsCl gradient as described by

10

15

20

25

30

Hull et al. Infect. Immun., 33: 933-938 (1981), which is incorporated herein by reference, and subsequently dialyzed against 4 changes of TE. BCG DNA was physically sheared by passage through a 22 gauge needle until an average size of 3-10 kb was obtained (20-25 passages). This DNA was then biotinylated using photobiotin (Clonetech, Palo Alto, California, USA) according to the method of Straus and Ausubel, Proc. Natl. Acad. Sci. USA, 87: 1889-1893 (1990), which is incorporated herein by reference.

DNA Subtraction

DNA subtraction was carried out between virulent *M. tuberculosis* H37Rv and avirulent BCG. H37R chromosomal DNA was selected because it was the most readily available chromosomal DNA from a virulent strain. In addition, *M. bovis* and *M tuberculosis* H37Rv are highly homologous.

M. bovis/M. tuberculosis specific probes were generated by the method of Straus and Ausubel, supra. with the following modifications. Sheared and biotinylated BCG DNA was used in a 10:1 excess for each round of subtraction. Wild type M. tuberculosis H37Rv DNA was digested with Sau3A to an average size of 1 kb. Hybridization conditions were 1M NaCl and 65 °C for 18 hours. Following five cycles (successive denaturation and reassociations) of subtraction, Sau3A1 adaptors (GACACTCTCGAGACATCACCGTCC and GATCGGACGGTGATGTCTCGAGAGTG were ligated to the subtraction product and amplified in a PCR reaction for 35 cycles (30 sec at 95°C, 30 sec at 55°C, and 3 min at 72°C). The M. tuberculosis/M. bovis specific probes were radiolabeled by using one strand of the adaptor (GACACTCTCGAGACATCACCGTCC) as a primer and labeling with ³²P dCTP using the Klenow fragment of DNA polymerase.

An M. bovis cosmid library was constructed in the BamH1 site of sCOS (Stratagene, La Jolla California, USA) with subsequent in vitro packaging and infection of E. coli strain NM554 (Stratagene). 600 colonies were picked to Nytran circular membranes and the membranes prepared according to the method of Grunstein and Hogness, Proc. Natl. Acad. Sci. USA, 72: 3961 (1975), which is incorporated herein by reference. These filters were then probed using the BCG subtracted probe and positive clones selected for further analysis. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly, Nucleic Acids. Res., 7: 1513 (1973) which is

WO 96/25519 PCT/US96/01938

37

incorporated herein by reference. Restriction fragments that hybridize with the MTB/MBV specific probe were further subcloned into pGEM7z or pGEM5z (Promega, Madison, Wisconsin, USA) for deletion analysis.

Plasmid DNA for DNA sequencing was prepared using Qiagen minicolumns (Qiagen Inc. Chatsworth California, USA) and sequenced by the method of Henikoff, *Gene*, 28: 351-359 (1984), which is incorporated herein by reference, using the Erase A Base System (Promega). DNA sequencing reactions were run using a Perkin Elmer 9600 thermocycler and analyzed on an automated ABI sequencer. Analysis and assembly of contiguous DNA sequence was done using the ABI analysis software and SeQuencher sequence analysis software by Gene Clones Corp (Ann Arbor, Michigan, USA).

Deletion Region 1 (BCGA1)

5

10

15

20

25

30

Sequence analysis of over 16 kb of MBV region 1 and homologous regions in BCG revealed the precise junctions for the deletion in BCG. Eight open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 4). The potential start and stop codons and predicted maximum protein coding capacity are listed in Figure 4. Consensus ribosomal binding site sequences were found near potential start codons for seven of eight open reading frames. TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for 3 of 8 open reading frames in region 1.

Most notable is the ORF1C homology to an unpublished and uncharacterized sequence listed in Genbank as M. tuberculosis antigen esat6. A 65 base pair repeated overlapping (repeated ~2 1/2 times) sequence was also recognized within the ORF1C (esat6) open reading frame. Also noteworthy are the significant homologies identified between ORF1H and bacterial serine proteases including B. subtilus subtilisin. Of the eight recognized open reading frames, four (ORFs 1B, 1C, 1D, and 1E) are located entirely within the 9 kb region deleted in BCG. One ORF traverses the BCG deletion junction in virulent M. bovis.

DNA probes from the 9 kb deletion in region 1 demonstrated that this region is absent in all BCG substrains and present in all virulent MBV and MTB strains tested. Furthermore, restriction fragment patterns observed in Southern blot analysis

10

15

20

25

30

with region 1 probes are non-polymorphic and identical in virulent MBV and MTB. This region has far fewer direct and indirect repeats than the regions 2 (BCG Δ 2) and 3 (BCG Δ 3) characterized below.

The sequence of a small region, estimated to be less than 20 bp between basepair coordinates 10654 and 10664 in region 1 has been recalcitrant to automated sequencing. Therefore, pending sequence confirmation, the base pair coordinates given in the region 1 map (Figure 4) are approximations. The precise sequence determination is likely to effect the Orf1E open reading frame.

Deletion Region 2 (BCGA2)

Sequence analysis of over 15 kb of MBV region 2 and homologous regions in BCG revealed the precise junctions for an 11 kb deletion in BCG. Thirteen open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 5). The potential start and stop codons and predicted maximum protein coding capacity are also shown in Figure 5. Candidate consensus sequences resembling ribosomal binding sites were found near potential start codons for eight open reading frames. Of the thirteen open reading frames recognized in BCGA2, nine are located entirely within the 11 kb region deleted in most BCG strains while ORF2B2 and ORF2I traverse the deletion junctions.

TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for five open reading frames in BCG Δ 2. A protein encoded by ORF2C exhibits striking similarity to the *E. coli* iciA protein which is thought to play a role in inhibiting and regulating the initiation of chromosomal replication. The iciA protein product is a member of the large LysR family of transcriptional regulatory proteins. Orf2F is highly homologous to an *S. typhimurium* ribonucleotide diphosphate reductase and a region of the *E. coli* and *S. typhimurium* proUVWX operon. Orf2H was found to have significant homology to *E. coli* and *S. typhimurium* permeases involved in aromatic amino acid transport and a eukaryotic cell retroviral receptor.

The Orf2G encoded protein was identical to the MTB mpt64 gene previously thought to encode a secreted antigen which is specifically expressed by MTB

39

and not BCG strains. Recent analysis of mpt64 expression revealed that three BCG substrains do express mpt64 (Moreau, Tokyo, Russian). Probes specific for mpt64 or other non-repetitive parts of region 2 hybridized to all MTB strains tested and the same three BCG substrains shown to express mpt64. Of interest is the finding that these three BCG substrains are derived from the original Pasteur strain prior to 1925. The current Pasteur strain and all strains derived from the original Pasteur strain after 1925, including the Connaught strain used in the subtractive analysis in this study, are deleted in the 11 kb DNA segment contained within BCG Δ 2. These data indicate that an additional mutational event deleting the 11 kb segment of region 2, occurred in the BCG Pasteur strain sometime after 1925.

Southern blot analysis with probes from different segments of region 2 revealed a repetitive element located within a 2 kb segment (8-10 kb) of region 2. This repetitive element is ubiquitous in all tubercle bacilli tested. This element provides a marker suitable for RFLP analysis of mycobacterial strains.

15

20

25

30

10

5

Deletion Region 3 (BCGA3)

Sequence analysis of the almost 11 kb region 3 sequence and comparison to a homologous region in BCG precisely identified the deletion junctions for BCG. Twelve potential open reading frames were recognized in the region 3 sequence, seven of which are entirely located within the 9 kb region deleted in BCG. At least 9 ORFs in BCGA3 exhibit codon usage preferences comparable to that of the tubercle bacilli. Sequence homology analysis of presumptive protein sequences encoded by six open reading frames in region 3 revealed highly significant homology to listed sequences. Orfs3B, 3D, and 3E exhibit homology to phage sequences, suggesting a phage derivation for 4 or more kb of DNA in region 3. Homology to putative open reading frames in two M. leprae cosmids was also observed including homology to a putative bid gene encoding a protein involved in biotin synthesis. Also of interest was homology between ORF3A and an MTB sequence (mce) associated with cell invasion and intracellular survival.

Southern blot analysis with segments of region 3 deleted in BCG revealed that prototype lab strains of virulent MBV and MTB all carry deletion region 3 DNA. However, clinical isolates from PHRI are highly polymorphic or deleted in region 3.

This region contains many large direct and indirect repeats and, as mentioned above, at least 2 ORFs are homologous to phage sequences including homology to DNA invertases or recombinases. The repetitive nature of this region and the possible presence of a DNA recombinase could explain the polymorphisms observed in this region.

5

The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, was recalcitrant to automated sequencing and remains to be determined. Therefore, the base pair coordinates given in the region 3 map (Figure 6) 3' to the 9kb marker are approximations. The precise sequence determination of region is likely to effect the length of open reading frames 3H and 3L.

10

15

The foregoing subtractive analysis identified 3 regions in virulent M. bovis and M. tuberculosis prototype strains which are deleted in the avirulent BCG strain. The deletion located in region 2 may not have arisen in the original BCG Pasteur strain as this region is only deleted in strains derived from the original Pasteur strain after 1925. Region 3 is present in virulent MTB and MBV lab prototype strains (H37Rv, Erdman) and is highly polymorphic and at least partially deleted in the majority of MTB clinical isolates tested. Region 1 is apparently conserved and intact in all virulent MBV and MTB strains tested to date while all avirulent BCG strains tested to date are missing approximately 9kb from region 1.

20

Example 2

Screening and Identification of an Avirulent Mycobacterium

The 32 P labeled subtraction probe obtained in Example 1, was used to probe EcoRI and BamHI restricted chromosomal DNAs from BCG Connaught, Mycobacterium bovis, and various strains of Mycobacterium tuberculosis in a Southern blot. The hybridization was performed at 70°C in 6X SSC overnight.

25

The resulting Southern blot is illustrated in Figure 8. The probe showed no labeling of BCG reflecting the presence of all three deletions, while the other strains were labeled.

30

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

WO 96/25519 PCT/US96/01938

WHAT IS CLAIMED IS:

A marker for an avirulent mycobacterium, said marker comprising 1. a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGala, BCGalab, BCGala BCGA3ab, BCGA1, BCGA2, and BCGA3.

The marker of claim 1, wherein said marker specifically hybridizes 2. under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG.

1 2

3

4

5

1

2

3

4

1

2

1

2

3

4

5

6

7

8

9

10

11

- The marker of claim 2, wherein said marker comprises a 3. subsequence of a nucleic acid where said nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGala, BCGalab, BCGalab, BCGalab, BCGalab, BCG△2ab, BCG△3ab, BCG△1, BCG△2, and BCG△3.
- 1 The marker of claim 2, wherein said marker is selected from the 4. group consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, 2 BCG△lab, BCG△2ab, BCG△3ab, BCG△1, BCG△2, and BCG△3. 3
 - The marker of claim 2, wherein said marker comprises a nucleic 5. acid having at least 90 percent sequence identity with a sequence selected from the group consisting of BCGala, BCGalb, BCGala, BCGalab, BCGalab, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3.
 - The marker of claim 2, wherein said marker comprises a 6. radioactive nucleotide probe.

1

4 5

1 The marker of claim 2, wherein said subsequence is a sequence 7. selected from an open reading frame of a deletion, said deletion being selected from the 2 group consisting of BCG\(\Delta\)1, BCG\(\Delta\)2, BCG\(\Delta\)3. 3 1 A polypeptide encoded by a subsequence of a deletion sequence 8. selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 The polypeptide of claim 8, wherein the subsequence is selected 1 9. from an open reading frame (ORF) of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 An antibody that binds specifically to the polypeptide of claim 8. 10. 1 A recombinant cell comprising a first nucleic acid that hybridizes 11. under stringent conditions with a second nucleic acid or a complement of said second 2 nucleic acid where said second nucleic acid or complement of said second nucleic acid is 3 selected from the group consisting of BCGala, BCGalb, BCGala, BCGalb, BCGala, 4 BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. 5 1 The recombinant cell of claim 11, wherein the cell is a 12. 2 Mycobacterium. 1 The cell of claim 11, wherein the cell expresses a polypeptide 13. encoded by an intact open reading frame from BCGa1, BCGa2, and BCGa3. 2 The cell of claim 11, wherein said cell is a mycobacterium having 14. one or more deletions in the genomic regions selected from the group consisting of 2 BCGa1, BCGa2, and BCGa3, wherein said deletions result in the attenuation of an 3 otherwise virulent strain of mycobacterium and wherein said deletions are present in up to two of said regions.

1	15. The mycobacterium of claim 14, wherein said deletions comprise a
2	deletion selected from the group consisting of BCGa1, BCGa2, and BCGa3.
1	16. A method of distinguishing between an attenuated and a virulent
2	mycobacterium, said method comprising detecting the presence or absence of a first
3	nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a
4	complement of said second nucleic acid where said second nucleic acid or complement of
5	said second nucleic acid is selected from the group consisting of BCGala, BCGalb,
6	BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1,
7	BCG∆2, and BCG∆3.
1	17. The method of claim 16, wherein said first nucleic acid specifically
2	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3	acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	10 77
2	18. The method of claim 17, wherein said first sequence is amplified
<u> </u>	prior to detection.
1	19. The method of claim 17, wherein said first sequence is amplified
2	by the polymerase chain reaction.
1	20. A method of claim 17, wherein said detecting comprises a Southern
2	blot.
1	
)	21. A method of claim 17, wherein said detecting comprises detecting a
۷	polypeptide encoded by said first nucleic acid.

WO 96/25519

6

BCG Δ 2, and BCG Δ 3.

	44
2	22. The method of claim 21, wherein the polypeptide is encoded by an
3	intact open reading frame of a nucleotide sequence selected from the group consisting of
4	BCGa1, BCGa2, and BCGa3.
1	23. The method of claim 21, wherein the polypeptide is visualized by
2	antibody hybridization.
1	24. A method for determining whether an attenuated or a virulent
2	Mycobacterium is present in a sample comprising:
3	providing a first nucleic acid that hybridizes under stringent conditions
4	with a second nucleic acid or a complement of said second nucleic acid where said
5	second nucleic acid or complement of said second nucleic acid is selected from the group
6	consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab,
7	BCG∆2ab, BCG∆3ab, BCG∆1, BCG∆2, and BCG∆3; and
8	hybridizing said first nucleic acid to the biological sample.
1	25. The method of claim 24, wherein said first nucleic acid specifically
2	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3	acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	26. A method of producing an attenuated Mycobacterium species, said
2	method comprising deleting from the genomic DNA of a virulent mycobacterium a first
3	nucleic acid that specifically hybridizes under stringent conditions with a second nucleic
4	acid or a complement of said second nucleic acid where said second nucleic acid or
5	complement of said second nucleic acid is selected from the group consisting of BCGa1,
	,

WO 96/25519 PCT/US96/01938

						17	63						
1 of 22	09	120	180	240	300	360	420	480	540	009	099	720	780
FIGURE 1-1 Page	GAATTCCTGC GCACCCTGAT CCTGTCGCTG GTGGCAATGA CTCATCCAGA TCAGGTGAAT	CTCCTGCTCA CCGACTTCAA AGGTGGTTCA ACCTTCCTGG GAATGGAAAA GCTTCCGCAC	ACTGCCGCTG TCGTCACCAA CATGGCCGAG GAAGCCGAGC TCGTCAGCCG GATGGGCGAG	GTGTTGACCG GAGAACTCGA TCGGCGCCAG TCGATCCTCC GACAGGCCGG GATGAAAGTC	GGCGCGCCG GAGCCCTGTC CGGCGTGGCC GAATACGAGA AGTACCGCGA ACGCGGTGCC	GACCTACCCC CGCTGCCAAC GCTTTTCGTC GTCGTCGACG AGTTCGCCGA GCTGTTGCAG	AGTCACCCGG ACTTCATCGG GCTGTTCGAC CGGATCTGCC GCGTCGGGCG GTCGCTGAGG	GTCCATCTGC TGCTGGCTAC CCAGTCGCTG CAGACCGGCG GTGTTCGCAT CGACAAACTG	GAGCCAAACC TGACATATCG AATCGCATTG CGCACCACCA GCTCTCATGA ATCCAAGGCG	GTAATCGGCA CACCGGAGGC GCAGTACATC ACCAACAAGG AGAGCGGTGT CGGGTTTCTC	CGGGTCGGCA TGGAAGACCC GGTCAAGTTC AGCACCTTCT ACATCAGTGG GCCATACATG	CCGCCGGCGG CAGGCGTCGA AACCAATGGT GAAGCCGGAG GGCCCGGTCA ACAGACCACT	AGACAAGCCG CGCGCATTCA CAGGTTCACC GCGGCACCGG TTCTCGAGGA GGCGCCGACA
	ď5	ົນ	AC	GJ	Ö	G.P.	AG	GI	GA	GI	S	S	AG

						2/6	3						
2 of 22	840	006	096	1020	1080	1140	1200	1260	1320	1380	1440	1500	1560
Page	BAGGAGCGG CGCCAACGGC	GCGCGCAT GACTGCTGAA	TCGGCACTGC TGAATCGCGT	AATCCGGTCC CGCTCAACGA GCTCATCGCC	GATGAACC GCGCCGCCAT	GCAACATCGG TATTGGGGGC	TGATGTCGGC CGCCGCCACA	GGCGCCG GCTGATCTAT	CCGAGCCCGA CAAGGTCAAC	AAACCACCTT CAAGGAACAC	CCAAGTCA ACCCGTTGCG	CCCGGTTT TGTCGGCGAG	GGGCTGGG GTTCGGCGTC
FIGURE 1-2	rgcaaagc gcagcgatga g	CGCAGCGA TGAGGAGGAG CC	CTGGACCAGC	CGTTGACC AATCCGGTCC CC	SATTTGCC CTGGGGATCA TO	ອວອອວວອອອອ	CAGACGATGG	rctattgc atcgacctag gi	FIGGGGTA GCCAATCGGT CC	CGGCAACGGG	ACCGGCAG CTGCGTGACG AT	TTCTGATC ATCGACGGAT GGCCCGGTTT	TICAAGAT CTGGCCGCCC AGGGGCTGGG
	CCGTGACCCG CGCCGGCGAC GATGCAAAGC GCAGCGATGA GGAGGAGCGG CGCCAACGGC	CCGCGCCGGC GACGATGCAA AGCGCAGCGA TGAGGAGGAG CGGCGCGCAT GACTGCTGAA	CCGGAAGTAC GGACGCTGCG CGAGGTTGTG	GCGTACAAGA TGTGGCTGCC GCCGTTGACC	CGTGATCGGC GACAACCCCT GCGATTTGCC CTGGGGATCA TGGATGAACC GCGCCGCCAT	CTACAGGATG TGTGGGGCGT AGACGTTTCC	GCACCTCAAA CCGGGAAGTC GACGCTACTG	CACTCACCGC GCAACGTTCA GTTCTATTGC ATCGACCTAG GTGGCGGCGG GCTGATCTAT	CTCGAAAACC TTCCACACGT CGGTGGGGTA GCCAATCGGT	CGGGTGGTCG CAGAGATGCA AGCCGTCATG	CGAGTGGGCT CGATCGGGAT GTACCGGCAG CTGCGTGACG ATCCAAGTCA ACCCGTTGCG	TCCGATCCAT ACGCCGACGT CTTTCTGATC	TTCCCCGACC TTGAGGGGCA GGTTCAAGAT

		FIGURE 1-3		Page 3	Page 3 of 22
CACGICATCA TCTCCACGCC ACGCTGGACA GAGCTGAAGT CGCGTGTTCG CGACTACCTC	ACGCTGGACA	GAGCTGAAGT	CGCGTGTTCG	CGACTACCTC	1620
GGCACCAAGA TCGAGTTCCG GCTTGGTGAC GTCAATGAAA CCCAGATCGA CCGGATTACC	SCTTGGTGAC	GTCAATGAAA	CCCAGATCGA	CCGGATTACC	1680
CGCGAGATCC CGGCGAATCG I	TCCGGGTCGG	GCAGTGTCGA	TGGAAAAGCA CCATCTGATG	CCATCTGATG	1740
ATCGGCGTGC CCAGGTTCGA C	CGGCGTGCAC	AGCGCCGATA	ACCTGGTGGA	GGCGATCACC	1800
GCGGGGGTGA CGCAGATCGC I	rtcccagcac	TTCCCAGCAC ACCGAACAGG	CACCTCCGGT GCGGGTCCTG	GCGGGTCCTG	1860
CCGGAGCGTA TCCACCTGCA C	CGAACTCGAC	CGAACTCGAC CCGAACCCGC CGGGACCAGA GTCCGACTAC	CGGGACCAGA	GTCCGACTAC	1920
CGCACTCGCT GGGAGATTCC G	GATCGGCTTG	CGCGAGACGG	CGCGAGACGG ACCTGACGCC GGCTCACTGC	GGCTCACTGC	1980
CACATGCACA CGAACCCGCA C	CCTACTGATC	TTCGGTGCGG	CCAAATCGGC CAAGACGACC	CAAGACGACC	2040
ATTGCCCACG CGATCGCGCG C	CGCCATTTGT	GCCCGAAACA GTCCCCAGCA GGTGCGGTTC	GTCCCCAGCA	GGTGCGGTTC	2100
ATGCTCGCGG ACTACCGCTC GGGCCTGCTG GACGCGGTGC CGGACACCCCA TCTGCTGGGC	GGCCTGCTG	GACGCGGTGC	CGGACACCCA	TCTGCTGGGC	2160
GCCGGCGCGA TCAACCGCAA C	CAGCGCGTCG	CTAGACGAGG	CTAGACGAGG CCGCTCAAGC ACTGGCGGTC	ACTGGCGGTC	2220
AACCTGAAGA AGCGGTTGCC G	GCCGACCGAC	CTGACGACGG	CGCAGCTACG	CTCGCGTTCG	2280
TGGTGGAGCG GATTTGACGT C	CGTGCTTCTG	CGTGCTTCTG GTCGACGATT GGCACATGCA GCCGTGGGTG	GGCACATGCA	GCCGTGGGTG	2340

FIGURE 1-4	4 of 22
CCGCCGGGGG GATGCCGCCG ATGGCACCGC TGGCCCCGTT ATTGCCGGCG GCGCAGATA	2400
TCGGGTTGCA CATCATTGTC ACCTGTCAGA TGAGCCAGGC TTACAAGGCA ACCATGGACA	2460
AGTTCGTCGG CGCCGCATTC GGGTCGGGCG CTCCGACAAT GTTCCTTTCG GGCGAGAAGC	2520
AGGAATTCCC ATCCAGTGAG TTCAAGGTCA AGCGGCGCCC CCCTGGCCAG GCATTTCTCG	2580
TCTCGCCAGA CGGCAAAGAG GTCATCCAGG CCCCCTACAT CGAGCCTCCA GAAGAAGTGT	2640
TCGCAGCACC CCCAAGCGCC GGTTAAGATT ATTTCATTGC CGGTGTAGCA GGACCCGAGC	2700
TCAGCCCGGT AATCGAGTTC GGGCAATGCT GACCATCGGG TTTGTTTCCG GCTATAACCG	2760
AACGGTTTGT GTACGGGATA CAAATACAGG GAGGGAAGAA GTAGGCAAAT GGAAAAAATG	2820
TCACATGATC CGATCGCTGC CGACATTGGC ACGCAAGTGA GCGACAACGC TCTGCACGGC	2880
GTGACGGCCG GCTCGACGGC GCTGACGTCG GTGACCGGGC TGGTTCCCGC GGGGCCCGAT	2940
GAGGTCTCCG CCCAAGCGGC GACGGCGTTC ACATCGGAGG GCATCCAATT GCTGGCTTCC	3000
AATGCATCGG CCCAAGACCA GCTCCACCGT GCGGGCGAAG CGGTCCAGGA CGTCGCCCGC	3060
ACCTATTCGC AAATCGACGA CGGCGCCGCC GGCGTCTTCG CCTAATAGGC CCCCAACACA	3120

		FIC	FIGURE 1-5		Page 5	of 22	
TCGGAGGGAG	TCGGAGGGAG TGATCACCAT GCTGTGGCAC GCAATGCCAC CGGAGCTAAA TACCGCACGG	GCTGTGGCAC	GCAATGCCAC	CGGAGCTAAA	TACCGCACGG	3180	
CTGATGGCCG	CTGATGGCCG GCGCGGGTCC GG	GGCTCCAATG	CTTGCGGCGG	CCGCGGGATG	GCAGACGCTT	3240	
TCGGCGGCTC	TCGGCGGCTC TGGACGCTCA GG	GGCCGTCGAG	TTGACCGCGC	CCGTCGAG TTGACCGCGC GCCTGAACTC TCTGGGAGAA	TCTGGGAGAA	3300	
GCCTGGACTG	GCCTGGACTG GAGGTGGCAG CGACAAGGCG CTTGCGGCTG CAACGCCGAT GGTGGTCTGG	CGACAAGGCG	CTTGCGGCTG	CAACGCCGAT	GGTGGTCTGG	3360	
CTACAAACCG	CTACAAACCG CGTCAACACA GG	GGCCAAGACC	CGTGCGATGC	AGGCGACGGC	GCAAGCCGCG	3420	
GCATACACCC	GCATACACCC AGGCCATGGC CACGACGCCG	CACGACGCCG	TCGCTGCCGG	AGATCGCCGC CAACCACATC	CAACCACATC	3480	5/6
ACCCAGGCCG	ACCCAGGCCG TCCTTACGGC CACCAACTTC TTCGGTATCA ACACGATCCC GATCGCGTTG	CACCAACTTC	TTCGGTATCA	ACACGATCCC	GATCGCGTTG	3540	3
ACCGAGATGG	ACCGAGATGG ATTATTTCAT CCGTATGTGG		AACCAGGCAG	AACCAGGCAG CCCTGGCAAT GGAGGTCTAC	GGAGGTCTAC	3600	
CAGGCCGAGA	CAGGCCGAGA CCGCGGTTAA CA	CGCTTTTC	GAGAAGCTCG	AGCCGATGGC	GTCGATCCTT	3660	
GATCCCGGCG	GATCCCGGCG CGAGCCAGAG CA	CACGACGAAC	CCGATCTTCG	CGACGAAC CCGATCTTCG GAATGCCCTC CCCTGGCAGC	CCCTGGCAGC	3720	
TCAACACCGG	TCAACACCGG TTGGCCAGTT GCCGCCGGCG GCTACCCAGA CCCTCGGCCA ACTGGGTGAG	ອນອອນນອນນອ	GCTACCCAGA	CCCTCGGCCA	ACTGGGTGAG	3780	
ATGAGCGGCC	ATGAGCGGCC CGATGCAGCA GCTGACCCAG	GCTGACCCAG	CCGCTGCAGC AGGTGACGTC		GTTGTTCAGC	3840	
CAGGTGGGCG	CAGGTGGGCG GCACCGGCGG CGGCAACCCA		GCCGACGAGG	AAGCCGCGCA	GATGGGCCTG	3900	

						6/6	3						
6 of 22	3960	4020	4080	4140	4200	4260	4320	4380	4440	4500	4560	4620	4680
Page (CAGCGCGGGC	CCGCACGCCG	CGGCTGCTGC	GCCAGGGTGC	CGCAGGAGCG	CCGTAATGAC	GGAAGGTAAA	CGCGCAGGAG	GGTGGAGTCG	CCAGGCCGCG	GATCTCGACG	GCAGGCGCTG	AACATGACAG
	GATCAGGCCC	GGTCGTTGAC	TGCCCCCTCG GTGATGCCGG CGGCTGCTGC	CGGATCGTCG GCGACGGGTG GCGCCGCTCC GGTGGGTGCG GGAGCGATGG GCCAGGGTGC	GCAATCCGGC GGCTCCACCA GGCCGGGTCT GGTCGCGCCG GCACCGCTCG CGCAGGAGCG	TGGTGAGCTC	AACAGACTTC CCGGCCACCC GGGCCGGAAG ACTTGCCAAC ATTTTGGCGA GGAAGGTAAA	GAGAGAAAGT AGTCCAGCAT GGCAGAGATG AAGACCGATG CCGCTACCCT CGCGCAGGAG	CTGAAAACCC AGATCGACCA GGTGGAGTCG	GGACGGCCGC	GTGGTGCGCT TCCAAGAAGC AGCCAATAAG CAGAAGCAGG AACTCGACGA GATCTCGACG	AATATTCGTC AGGCCGGCGT CCAATACTCG AGGGCCGACG AGGAGCAGCA GCAGGCGCTG	ACGGAGCAAA AACATGACAG
FIGURE 1-6	CTGGCTGGTG	GGCGCAGGTG	TGCCCCCTCG	GGTGGGTGCG	GGTCGCGCCG	AGAGGACGAC	ACTTGCCAAC	AAGACCGATG	CTGAAAACCC	9909909099	CAGAAGCAGG	AGGGCCGACG	ACGAAAAGAA
[FI]	GAACCATCCG	GTCGCTACCT	AAAGCCGGTT	CCCCCCTCC	GGCCGGGTCT	ACTGGGACGA	GGGCCGGAAG	GGCAGAGATG	CTCCGGCGAC	CCAGTGGCGC	AGCCAATAAG	CCAATACTCG	ACCCGCTAAT
	CTCGGCACCA GTCCGCTGTC GAACCATCCG	GCGGGCCTGC TGCGCGCGGA GTCGCTACCT	CTGATGTCTC AGCTGATCGA AAAGCCGGTT	GCGACGGGTG	GGCTCCACCA	TGAAGAAGAC GACGAGGACG ACTGGGACGA	CCGGCCACCC	AGTCCAGCAT	GCAGGTAATT TCGAGCGGAT CTCCGGCGAC	ACGGCAGGTT CGTTGCAGGG CCAGTGGCGC	TCCAAGAAGC	AGGCCGGCGT	TCCTCGCAAA TGGGCTTCTG ACCCGCTAAT
	CTCGGCACCA	GCGGGCCTGC	CTGATGTCTC	CGGATCGTCG	GCAATCCGGC	TGAAGAAGAC	AACAGACTTC	GAGAGAAAGT	GCAGGTAATT	ACGGCAGGTT	GTGGTGCGCT	AATATTCGTC	TCCTCGCAAA

						7/6	3						
of 22	4740	4800	4860	4920	4980	5040	5100	5160	5220	5280	5340	5400	5460
Page 7	GGAAATGTCA	GACCAAGCTC GCAGCGGCCT	GGGGCGGTAG CGGTTCGGAG GCGTACCAGG GTGTCCAGCA AAAATGGGAC GCCACGGCTA	CAGAACCTGG CGCGGACGAT CAGCGAAGCC GGTCAGGCAA	GAGTTCGCGT	TCGGTCTCGC CCTTTCTCGT	ACAAGCTCTT	TCTTCGACCC	ACGCCCAGAC	CGCCGCCACC	AGCCGCCCTC	GACCCGAACC	CGGCCCCACC
	CGCAATCCAG	GACCAAGCTC	AAAATGGGAC	CAGCGAAGCC	GGGCAACGCC		GCCGACTACG	GCGCAGCCGT	CCGAAGCCCA	TCGGCCCCGC	GCCGCAGGAG	CCCATCGCCG	GGACCCGAAC
FIGURE 1-7	CCGCGGCAAG	AGCAGTCCCT	GTGTCCAGCA	CGCGGACGAT	TGTTCGCATA	CGACCTTCCG	TGTCATGGCG	CGATATGGCA GCGCAGCCGT	GGCAAACCTA	GCGGTTCGTG	GATGCCGATC	ACCCCCCATG	GCCCATCGCC
FIC	GGTATCGAGG	GACGAGGGGA	GCGTACCAGG	CAGAACCTGG	GTCACTGGGA	CGGGCGAGTT	TCTGAGAGGT			ACCTGTCGGA	CGCCAACTCC	AACCACCCAC	CACCCCCCAT
	AGCAGCAGTG GAATTTCGCG GGTATCGAGG CCGCGGCAAG CGCAATCCAG	CGTCCATTCA TTCCCTCCTT	CGGTTCGGAG	CAACGCGCTG	TGGCTTCGAC CGAAGGCAAC GT	AACACGGGAT	GTTTATACGT TTGAGCGCAC TCTGAGAGGT TGTCATGGCG GCCGACTACG ACAAGCTCTT	CCGGCCGCAC GAAGGTATGG AAGCTCCGGA	CAGTGCTTCG TTTCCGCCGG CGCCCGCATC	TCCGCCCCCG ACGTCCGACG ACCTGTCGGA GCGGTTCGTG TCGGCCCCCGC CGCCGCCACC	CCCACCCCCA CCTCCGCCTC CGCCAACTCC GATGCCGATC GCCGCAGGAG AGCCGCCCTC	GCCGGAACCG GCCGCATCTA AACCACCCAC	GGCCCCACCC AAACCACCCC ACCCCCCAT GCCCATCGCC GGACCCGAAC CGGCCCCACC
	AGCAGCAGTG	CGTCCATTCA	GGGGCGGTAG	CCGAGCTGAA CAACGCGCTG	TGGCTTCGAC	AGAATAGCGA AACACGGGAT	GTTTATACGT	CCGGCCGCAC	CAGTGCTTCG	TCCGCCCCCG	CCCACCCCCA	GCCGGAACCG	GGCCCCACCC

8
o£
ω
Page
ෙත
∞ -
E E
GURE
FIGURE 1

FIGURE 1-8	Page 8 of 22
CAAACCACCC ACACCTCCGA TGCCCATCGC CGGACCTGCA CCCACCCCAA CCGAATCCCA	rccca 5520
GITGGCGCCC CCCAGACCAC CGACACCACA AACGCCAACC GGAGCGCCGC AGCAACCGGA	CCGGA 5580
ATCACCGGCG CCCCACGTAC CCTCGCACGG GCCACATCAA CCCCGGCGCA CCGCACCAGC	CCAGC 5640
ACCECCTGG GCAAAGATGC CAATCGGCGA ACCCCCGCCC GCTCCGTCCA GACCGTCTGC	rcrgc 5700
GTCCCCGGCC GAACCACCGA CCCGGCCTGC CCCCCAACAC TCCCGACGTG CGCGCCGGGG	3GGGG 5760
TCACCGCTAT CGCACAGACA CCGAACGAAA CGTCGGGAAG GTAGCAACTG GTCCATCCAT	CCAT 5820
CCAGGCGCGG CTGCGGGCAG AGGAAGCATC CGGCGCGCAG CTCGCCCCCG GAACGGAGCC	BAGCC 5880
CTCGCCAGCG CCGTTGGGCC AACCGAGATC GTATCTGGCT CCGCCCACCC GCCCCGCGCC	3CGCC 5940
GACAGAACCT CCCCCCAGCC CCTCGCCGCA GCGCAACTCC GGTCGGCGTG CCGAGCGACG	CGACG 6000
CGICCGACCC CGATITAGCC GCCCAACATG CCGCGGCGCA ACCTGATICA ATTACGGCCG	0909 52255
CAACCCACTG GCGGTCGTCG CCGCAAGCGT GCAGCGCCGG GATGCTCGAC GCGACACAAG	CAAG 6120
AAATCCTTAA GGCCGGCGGC CAAGGGGCCG AAGGTGAAGA AGGTGAAGCC CCAGAAACCG	ACCG 6180
AAGGCCACGA AGCCGCCCAA AGTGGTGTCG CAGCGCGGCT GGCGACATTG GGTGCATGCG	TGCG 6240

0
σ
Page
6-
FIGURE 1
FIG

	Y.	FIGURE 1-9		Page 9 of 22	of 22
TTGACGCGAA TCAACCTGGG	CCTGTCACCC	GACGAGAAGT	CCTGTCACCC GACGAGAAGT ACGAGCTGGA CCTGCACGCT	CCTGCACGCT	6300
CGAGTCCGCC GCAATCCCCG	CGGGTCGTAT	CAGATCGCCG	CGGGTCGTAT CAGATCGCCG TCGTCGGTCT CAAAGGTGGG	CAAAGGTGGG	6360
GCTGGCAAAA CCACGCTGAC	AGCAGCGTTG	GGGTCGACGT	TGGCTCAGGT	GCGGCCGAC	6420
CGGATCCTGG CTCTAGACGC	GGATCCAGGC	GCCGGAAACC	TCGCCGATCG	GGTAGGGCGA	6480
CAATCGGGCG CGACCATCGC	TGATGTGCTT	GCAGAAAAAG	AGCTGTCGCA	CTACAACGAC	6540
ATCCGCGCAC ACACTAGCGT	CAATGCGGTC	AATCTGGAAG	CAATGCGGTC AATCTGGAAG TGCTGCCGGC ACCGGAATAC	ACCGGAATAC	0099
AGCTCGGCGC AGCGCGCGCT	CAGCGACGCC	GACTGGCATT	CAGCGACGCC GACTGGCATT TCATCGCCGA TCCTGCGTCG	TCCTGCGTCG	0999
AGGTTTTACA ACCTCGTCTT	GGCTGATTGT	GGGGCCGGCT	GGGGCCGGCT TCTTCGACCC	GCTGACCCGC	6720
GGCGTGCTGT CCACGGTGTC	CGGTGTCGTG	GTCGTGGCAA	GTGTCTCAAT	CGACGGCGCA	6780
CAACAGGCGT CGGTCGCGTT	GGACTGGTTG	CGCAACAACG	CGCAACAACG GTTACCAAGA TTTGGCGAGC	TTTGGCGAGC	6840
CGCGCATGCG TGGTCATCAA	TCACATCATG		CCGGGAGAAC CCAATGTCGC AGTTAAAGAC	AGTTAAAGAC	0069
CTGGTGCGGC ATTTCGAACA	GCAAGTTCAA	ອອອລລອອລລລ	TCGTGGTCAT GCCGTGGGAC	GCCGTGGGAC	0969
AGGCACATTG CGGCCGGAAC	CGAGATTTCA	CTCGACTTGC	CTCGACTTGC TCGACCCTAT	CTACAAGCGC	7020

FIGURE 1-10	Page 10 of 22
AAGGTCCTCG AATTGGCCGC AGCGCTATCC GACGATTTCG AGAGGGCTGG ACGTCGTTGA	3A 7080
GCGCACCTGC TGTTGCTGCT GGTCCTACCG CCGCGGGGC AACCGCTGCG CGGCCTGCCA	CA 7140
CCACCCGGGT GACGATCCTG ACCGGCAGAC GGATGACCGA TTTGGTACTG CCAGCGGCGG	3G 7200
TGCCGATGGA AACTTATATT GACGACACCG TCGCGGTGCT TTCCGAGGTG TTGGAAGACA	ZA 7260
CGCCGGCTGA TGTACTCGGC GGCTTCGACT TTACCGCGCA AGGCGTGTGG GCGTTCGCTC	rc 7320
GTCCCGGATC GCCGCCGCTG AAGCTCGACC AGTCACTCGA TGACGCCGGG GTGGTCGACG	.c 7380
GGTCACTGCT GACTCTGGTG TCAGTCAGTC GCACCGAGCG CTACCGACCG TTGGTCGAGG	3G 7440
ATGTCATCGA CGCGATCGCC GTGCTTGACG AGTCACCTGA GTTCGACCGC ACGGCATTGA	3A 7500
ATCGCTTTGT GGGGGGGGGG ATCCCGCTTT TGACCGCGCC CGTCATCGGG ATGGCGATGC	1560
GGGCGTGGTG GGAAACTGGG CGTAGCTTGT GGTGGCCGTT GGCGATTGGC ATCCTGGGGA	A 7620
TCGCTGTGCT GGTAGGCAGC TTCGTCGCGA ACAGGTTCTA CCAGAGCGGC CACCTGGCCG	.G 7680
AGTGCCTACT GGTCACGACG TATCTGCTGA TCGCAACCGC CGCAGCGCTG GCCGTGCCGT	T 7740
TGCCGCGCGG GGTCAACTCG TTGGGGGCGC CACAAGTTGC CGGCGCCGCT ACGGCCGTGC	12 7800

		F	FIGURE 1-11		Page 1	11 of 22	
TGTTTTTGAC	TGTTTTTGAC CTTGATGACG CGGGGGGGCC	റാളൊള്ളോ	CTCGGAAGCG	TCATGAGTTG	GCGTCGTTTG	7860	
CCGTGATCAC	CCGTGATCAC CGCTATCGCG GTCATCGCGG	GTCATCGCGG	CCGCCGCTGC	CTTCGGCTAT GGATACCAGG	GGATACCAGG	7920	
ACTGGGTCCC	ACTGGGTCCC CGCGGGGGG ATCGCATTCG GGCTGTTCAT TGTGACGAAT GCGGCCAAGC	ATCGCATTCG	GGCTGTTCAT	TGTGACGAAT	GCGGCCAAGC	7980	
TGACCGTCGC	TGACCGTCGC GGTCGCGCGG ATCGCGCTGC CGCCGATTCC GGTACCCGGC GAAACCGTGG	ATCGCGCTGC	CGCCGATTCC	GGTACCCGGC	GAAACCGTGG	8040	
ACAACGAGGA	ACAACGAGGA GTTGCTCGAT CCCGTCGCGA	CCCGTCGCGA	CCCCGGAGGC	TACCAGCGAA GAAACCCCGA	GAAACCCCGA	8100	
CCTGGCAGGC	ccreecaege carcareece reserecee	TCGGTGCCCG		CGTCCGCGGT CCGGCTCACC GAGCGCAGCA	GAGCGCAGCA	8160	11/6
AACTGGCCAA	AACTGGCCAA GCAACTTCTC ATCGGATACG TCACGTCGGG CACCCTGATT CTGGCTGCCG	ATCGGATACG	TCACGTCGGG	CACCCTGATT	CTGGCTGCCG	8220	3
GTGCCATCGC	GTGCCATCGC GGTCGTGGTG CGCGGGCACT	CGCGGGCACT	TCTTTGTACA	TCTTTGTACA CAGCCTGGTG	GTCGCGGGTT	8280	
TGATCACGAC	TGATCACGAC CGTCTGCGGA TTTCGCTCGC	TTTCGCTCGC	GGCTTTACGC	CGAGCGCTGG	TGTGCGTGGG	8340	
CGTTGCTGGC	CGTTGCTGGC GGCGACGGTC GCGATTCCGA CGGGTCTGAC GGCCAAACTC ATCATCTGGT	GCGATTCCGA	CGGGTCTGAC	GGCCAAACTC	ATCATCTGGT	8400	
ACCCGCACTA	ACCCGCACTA TGCCTGGCTG TTGTTGAGCG	TTGTTGAGCG		TCTACCTCAC GGTAGCCCTG GTTGCGCTCG	GTTGCGCTCG	8460	
TGGTGGTCGG	TGGTGGTCGG GTCGATGGCT CACGTCCGGC	CACGTCCGGC	GCGTTTCACC	GGTCGTAAAA	CGAACTCTGG	8520	
AATTGATCGA	AATTGATCGA CGGCGCCATG ATCGCTGCCA TCATTCCCAT GCTGCTGTGG ATCACCGGGG	ATCGCTGCCA	TCATTCCCAT	GCTGCTGTGG	ATCACCGGGG	8580	

					12/	63						
8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
CTGATTGG CGGTTCCTGA	CCCGACAA ATTGCTGCGA	CACGTATA GGAGATCCGG	CAGCGGCC GCGAAATTGG	GCGGAACG GATTCGGTGG	TCAGTGAC GGGCTGCCCG	CGGCGCG GACGTCTATG	CATTCGGC TCGTCGGGCG	AGGCTACC CAGCTGCTGA	CCGCTGAG CTGGCACCCC	ACGCCGTT CAGATGTCGC	AACAGGCC GCCCAGAGCG	CGCTGAAAAA CCGGCCACCG
		AA	5) T	CA	BB -	G	IG		GG.	GC	CC	
GAGCCGGATC		CTCGATGGGG	CCGGCTTGAG	CGATCGCGGT		CCAACATGAA		GTCAGCCAAG		AGCTGGCTCC	GTCAAACCGC	AGCTTGCCAG
ATCCGGTTCT	GCAGGTTTGC	GTAAAATTTG	GTCGATCCCA	CCTCCGGCGC	ATGCCAAGCA	CGAACAGCAT	GGAACCAGTT	TCGGTCGGTG	ACGACCCAGC	CAACTCGTTC	CAGACGATCA	ATGCCCGCAC
GGTCCGCAAT	AGGACACGGC	GGCGCGTCCG	ACCGTTGGCC	TTTTCCGCAG	CAACAAGACC	CGCCCTGACT	TCAGTCACTG	TGGCGTCGCC	GTCACAGGTC	GACGGTGCCG	CCCCATCGCT	CGCAGGGCGG CAGCGGCCCA ATGCCCGCAC
TGTACGACAC	CAGAACATCG	TTGAGCGTGT	CAATGGCTGA	CCGGCCTCGT	TAGCAGCAAT	GCGTGAAAGC	CGAAGACCGA	AAGGCCTGGC	GCACACCCGT	GTGTTGTTGC	AAAACGCATC	CGCAGGGCGG
		CCGGTTCT GAGCCGGATC GGCTGATTGG CGGTTCCTGA AGGTTTGC ATACCTTCGG CGCCCGACAA ATTGCTGCGA	CCGGTTCT GAGCCGGATC GGCTGATTGG CGGTTCCTGA AGGTTTGC ATACCTTCGG CGCCCGACAA ATTGCTGCGA			8640 8700 8760 8820 8880	GAGCCGGATC GGCTGATTGG CGGTTCCTGA ATACCTTCGG CGCCCGACAA ATTGCTGCGA CTCGATGGGG AACACGTATA GGAGATCCGG CCGGCTTGAG CGCAGCGGCC GCGAAATTGG CGATCGCGGT CAGCGGAACG GATTCGGTGG TCGAATCGCT GGTCAGTGAC GGGCTGCCCG CCAACATGAA CGCGGCGG GACGTCTATG	8640 8700 8760 8820 8880 8940 9000	8640 8700 8760 8820 8940 9000	GAGCCGGATCGGCTGATTGGCGGTTCCTGAATACCTTCGGCGCCCGACAAATTGCTGCGACTCGATGGGGAACACGTATAGGAGGATCCGGCCGGCTTGAGCGCAGCGGCCGCGAAATTGGCCGGCTTGAGCGCAGCGGCCGGTTCGGTGGTCGAATCGCTGGTCAGTGACGGCTGCCCGCCAACATGAACGCGGCGGCGGACGTTATGCCAACATGAACGCGGCGGCGGACGTCTATGTGAGCCAGTATGCATTCGGCTCGTCGGCGTCAGCCAGAGTCAGGCTGACTCGTCGGCGTCGGCGAGACCAGCTGCTGA9180	GAGCCGGATCGGCTGATTGG8640ATACCTTCGGCGCCCCGACAAATTGCTGCGA8700CTCGATGGGGAACACGTATAGGAGATCCGG8760CCGGCTTGAGCGCAGCGGCCGCGAAATTGG8880CCGATCGCGGTCAGCGGAACGGATTCGGTGG8940TCGAATCGCTGGTCAGTGACGGCTGCTCCG9000TCGAATCGCTGGTCAGTGACGACGTCTATG9060TGAGCCAGTATGCATTCGGCTCGTCGGCGC9120TCGGCGAGCGCCGCTGAGCTGGCCACCC9180AGCTGGCTCCGCACGCCGTTCAGATGTCGC9240	8640 8700 8760 8820 8940 9000 9060 9120 9180 9240

						14/63	3						
14 of 22	10200	10260	10320	10380	10440	10500	10560	10620	10680	10740	10800	10860	10920
FIGURE 1-14 Page	GCCCAGGCTT GCACGGCCAA AACCGGGTAG GTGGCACAGC GTGCAATTTC GTCAACCGGG	ATTGCGTGAT CCGCGCTGGC CAAGTACACC TTATTCGGCA ATTCCATCCC GTCGGGTATG	TAGGCCAGCC CATAGCTGTT GGCCACGACG ATGGAACCGT CGGTGGTCAC CGCGGTGATC	CAGAAGAACC CGTAGTCGCC CGCGTTGTTG TCGGACGCGT TGAGCGCCCGC CGCGATGCGT	CGCGCCAACC GCAGCGCATC ACCGCGGCCA CGCTGGCGGG CGCTGGCAGC TGCAGTGGCG	GCGTCGCGTG CCGCCCGAGC CGCCGACACC GGGATCATCG ACACCGGCGT ACCGTCATCT	GCAGACTCGC TGCGATCGGG TTTGTCGATG TGATCGGTCG ACGGAGGGCG GGCAGGAGGT	GCCGTCCGCG CCGAGGCCGC CCGCGTGCTC GGTGCCGCCG CCTTGTCCGA GGTAGCCACC	TGCGTCCGCC CAGTGGCAGT ATGCGGACCC CGGAAAAAA AAACTCGAGT GCGTTCTTCG	GAGGTTTCCA ATTCTTGGAT TCCAGCACCC GCTCAGCGGT CTCGGCGACC AGACTGACAT	TGGCCCCATG CGTCGCCGTG ACCAATGAAT TGATGGCGGT ATGGCGCTCA TCAGCATCCA	GGCTAGAGTC ATTCTCCAGG ATATCGATCT CCCGTTGAGC GCCATCCACA TTATTGCCGA	TATCGGATTT AGCTTGCTCA ATCAACCCGG CAATATGCCT GTGCCAGGTA ATCACCGTGG

7						137 0	J						
15 of 22	10980	11040	11100	11160	11220	11280	11340	11400	11460	11520	11580	11640	11700
Page	CCGTTGGCAG	TGCTGGCGGC	ACCCTTTGCA AAACCTGGCT ATATTCCTGG GCCCGGTCAT	TCCACCCAGC CGCCCGGATC CAGCATCTGT CTGGCATAGC	CCAAACCGCC AGATCGCCTC	GGATCCACCC	TTGGGCGATC	CGGTGGTAAC	CAATATTGGC	GACAATGCGT	GTGCGTAGTC	ACATGGCCGA	CGGCCTGCGC
	ACCCAGGGCG	GTGGCCTTTC	ATATTCCTGG	CAGCATCTGT		CTCGGCCGCT	GATGCACCGC	ACGGGGCGCT	CTAGCTGCGG	TCAAAACATT	CGCTCATTCG	ATCGCGCGGG	GICGITALAC GCGGACGCCG CGGCCTGCGC
FIGURE 1-15	TGATGTTTGC	CTTCGAAGAC	AAACCTGGCT	CGCCCGGATC	CTCATCCCCT ACTGCCCTCC	TTCACGCCGG	CCCGTTGAAT CCGCGCGCAT GATGCACCGC TTGGGCGATC	TGCGCTGGCC GCGCTGTCGC ACGGGGCGCT CGGTGGTAAC	CTAAGACCAG	CCGATCCATG	SCTGAATGAC CGCATTGCGG CGCTCATTCG GTGCGTAGTC	CTTAGGCCAT TCCTTCGTTC ATCGCGCGG ACATGGCCGA	
F	ATCAATTGAT	CATAGGCCGC	ACCCTTTGCA	TCCACCCAGC	CTCATCCCCT	CCTCCGGCAT	CCCGTTGAAT	TGCGCTGGCC	GTAACCGAAC	CCGAACCCGG	GCTGAATGAC	CTTAGGCCAT	ລອລລອລລລອລ
	CGAGATAATC CTGCAGCGTC	CATTGGCGGC GCCGCCGGAC	AGGTGTCCAA TACATCGGTG	AGAAAGTGTC TTCATCGGCT	TGCCCGTCGG CCTGGTAATA	GCGGATCACC GTCCGGTTGG	CGCGCCGGTA TTCGCAGTAA	AGCCGGGTGG TCACCTCGCT	GGACGTCATA ATTAACCAGC	GACCAGGACT ATGGCGCCCT	ACTCACGCCG TGTCGGGCGC	GCTACCACCG CAACAATGGG	TAACGCAGCG GTCAGCTGCT
	CGAGATAATC	CATTGGCGGC	AGGTGTCCAA	AGAAAGTGTC	TGCCCGTCGG	GCGGATCACC	CGCGCCGGTA	AGCCGGGTGG	GGACGTCATA	GACCAGGACT	ACTCACGCCG	GCTACCACCG	TAACGCAGCG

22

FIGURE 1-16 CCCGCTGAGC CGCCGCCTCG GCACCCAGCT TCTTCAGCAA CGGTGAGCCA CTGGTGCCCA TTGATCGTCA CTTCGACGGT GGAAGGATCC GTTGTTCATC TGATTGAGCG TCCCGTCTAG CCAGCGTCAA CGCCCGGGCG ACATGCGGGT CCAATTCGTC

ACCGATTCGG CGCCCCCGAT	CGCGGATCCC	CGCGGATCCC AACGGCGCCG ACGGCACCCC GCCGCCTCCA	ACGGCACCCC	GCCGCCTCCA	12240
CCGCCACCGA GCGATGCCGC	TTTGACCGCC	TTTGACCGCC ACGTCGCCCG ACAGCGCTGC GGCTTCCCGC	ACAGCGCTGC	GGCTTCCCGC	12300
CCAGCCGACG TCAGCTGCGC	CGCCGTGTCA	CGCCGTGTCA GCCGGGAGGC CACCACCCGG CGATCCGGTA	CACCACCCGG	CGATCCGGTA	12360
GGCGGAACCA TCGGTGCGGC	TGGCATCCCG	TGGCATCCCG GTACCGGGAG TCACACCGGA GCCGTCAGAC	TCACACCGGA	GCCGTCAGAC	12420
GGCGGCATCA GGAAGCCAGG	GATCAATCCC	GATCAATCCC TGCTCTTGCG GAGGCGGGGC GGGTCGATCT	GAGGCGGGGC	GGGTCGATCT	12480

16/63

12180

GCGCCGCCGC CGGCCCTTCC CTGGCCTAAG CCGGCAATGT CACCAGCGCC AGCGGGCCGC

						17/6	63						
17 of 22	12540	12600	12660	12720	12780	12840	12900	12960	13020	13080	13140	13200	13260
Page	ACCGGTTCCA GGGCTGCCTT GTTGTTGTAT	ACACCGGGAG AATTTGGTCG	PATGTCTTC ATAAGTCGGA	GAGCCTGCTT GGCCATCGCA	GATCGAGCGA AGCCTCGCAA	ACCGCTTGAC GTCGCCTTGC	BAGCGATGC GCCTTGGTCG	STTGGGTTC ACCGGCCGTG	rccgacggc cccggccgaT	TCGCAGCCTC CTCATCAACC	ACGCTGCCG CTCTTTGGCA	CAGCTGTTG GGCGGCGTTT	TCGGTGGGTC CGCCATCGGG
FIGURE 1-17	GCGGGTTT	TCTGCTGA TACTCCGCGT	TAAAGCCGT TCGAGCCCGA C		TGCCGGTGTT		CACCCATCC GCAAAGTGCG C	CGCTTCTTTG AGATCCATGA AGTTGGGTTC ACCGGCCGTG	AGTICGGCC GAACTGICCC CICCGACGGC	TCCAGCGCGG	TTGCGCAGC GAGGTCGCCA G	ATGTTGTCG GCGGACAATA C	GGGACATCAG
	TGATGGCGGG GGGGAGGCTT CG	TCGGTCAGCA CCTTCTCCGA CC	CGGGCCGAAG GGTTTTCCGC GTAAAGCCGT TCGAGCCCGA CTATGTCTTC ATAAGTCGGA	TGTTCCCGCC TAGCCCACAC GTGCAGCTGC GCGACATATT	GCGCTCAATT TGGCCATGTG GAGTATCCAT	GCGGTAGCCG CATCGCCTTC CCAGTTGTCA AACCCCCGGA	AGCGTCAGGT TGAAAGTGTT CCACCCATCC GCAAAGTGCG CGAGCGATGC GCCTTGGTCG	CCCGTTTCGA GCTTCCTTGC C	GCCACCCTCG GCGTATCGGT TAGTTCGGCC	TCTGCCTGCA CAGTTCCTTC GCCGTCGTTG	TCGCCATACG CCTTGGCCGC GTTGCGCAGC GAGGTCGCCA GACGCTGCCG CTCTTTGGCA	CCGGCCGCCA GGTATTCCCG CATGTTGTCG GCGGACAATA CCAGCTGTTG	TTAGCCGCCG TGAGTTCGCA CGGTGTGATG

						18/6	3						
Page 18 of 22	13320	13380	13440	13500	13560	13620	13680	13740	13800	13860	13920	13980	14040
Page	GCCTCCACCT CGTTGGCCCT GTTCAAAATC TCTTGCTGAT CCACCGTCAC GGTCTGCGAC	A AACTGAAAGG	r cecctactt	CGGCAGCGCC ATCTGGTAGC GGCTTTCCTC GGGTGGGGAA ACCCGGCGAA TCGGCAGCTG	CCGCAGAATC ACCCGGTCAA TACCGGGATG	CGATGTGTGC	S CACCTGCCGC	CGCCGCACGC ACCACGAACT GGGTGAATGT CTGAGCGTCA CCCAGGTTGA GGGCGATGTC	TCGGTTCACC GTCTCGCCGA CCAGTACCCC	CTTGCAGCGC	GGGCCGCTGC CCGCCAAATA GGCGGGCGAA GCCCCTGGGT GTCTTGGGCT TGTCCGCCGT	CCCCGGCGCG ACCCGGACTC TGGTGATGGT	GGCGCCGCGT AGGCGGCAGT
	CCACCGTCA	ATCGTCGCTA	GCGAACGGAT	ACCCGGCGA	ACCCGGTCA	ACCTCCTGCC	GTGACCATCG	CCCAGGTTGA	GTCTCGCCG1	AGATGCTGGC	GTCTTGGGC	ACCCGGACTC	
FIGURE 1-18	TCTTGCTGAT	TATAGCCATT	TGCCGGCATC	GGGTGGGGAA	CCGCAGAATC	CGCCACCTTT	TGGCCCGACT	CTGAGCGTCA	TCGGTTCACC	GTTGGCCACC	GCCCTGGGT	ອນອນອອນນນ	CGGACCTCCG
FIC	GTTCAAAATC	TCCTTAGTGC		GGCTTTCCTC	TCACATTGTG		CCTCGAACTG TGGCCCGACT	GGGTGAATGT			GGCGGCGAA	GCGGGGCCAT	
	CGTTGGCCCT	TGCGTCATAT CGGATCATCC TCCTTAGTGC	TTCCTGCACT AATTTGATGC CGCCCGTTCA	ATCTGGTAGC	CCGATGCCGC GGGGTACCGA TCACATTGTG	CGGGCCGAGA TAGGTCGTCG CATTCGGCCA		ACCACGAACT	GACATCGTCG AAGGGCATGT AGACCGGGCA	AGCTGACCCG ATCGGCAGCT GGCAGTGGCG	CCGCCAAATA	GGTCAGCAAC ACCGTGGACT GCGGGGCCAT	GTGGTCCGCG CGCCCCACC ACCATACATC
	GCCTCCACCT	TGCGTCATAT	TTCCTGCACT	CGGCAGCGCC	CCGATGCCGC	CGGGCCGAGA	GCCGATCAAC CGGGCAAATT	CGCCGCACGC	GACATCGTCG	AGCTGACCCG	GGGCCGCTGC	GGTCAGCAAC	GTGGTCCGCG

	FIC	FIGURE 1-19		Page	19 of 22	
GTAGGCATCG CGCCCCTTGA TCATCGACCA TTTCTCCCGC ACAAAGCCGA TGTCGGTGGC	TCATCGACCA	TTTCTCCCGC	ACAAAGCCGA	TGTCGGTGGC	14100	
GTGGTCGTAG TCATCGAAGC TG	TGCGGCCACA	CACCGCGTCG	CGGCCACA CACCGCGTCG ACACCATGGC TAGCCAGTCG	TAGCCAGTCG	14160	
ATCGGCAATG CGCGTCGCGG ACGCCACCAA	ACGCCACCAA	ATACCGGGCC	AGTCCTGCGA	CGCCTTCATC	14220	
GCGGCGCTGC GCCGATTTGC GGGTGCGTTC CGGGTCGGCG	GGGTGCGTTC	CGGGTCGGCG	CGCAGCACGA	TCCAGGTCCG	14280	
GCGGTTCGCC GGCGCCGGGT	CTGTCCCGAT	CACCTGCTGA	CTGTCCCGAT CACCTGCTGA TACAGACTCA CCACGTCCGG	CCACGTCCGG	14340	
CGCTGCGGTA TTGCCGACGC GGTAGCCGGC	GGTAGCCGGC	TGAGACGATA	TCGGCCTCCA	AGTCGGGACA	14400	19/6
GTGCACCGAC AGGAGCTCCT CCACCAGTCC	CCACCAGTCC	GGTGTCCAGC	ATGTCGTCGG	TGTGGGCTTG	14460	3
CCCGTCGACG ATGACCGTCG GCGTGAATGG TCGGGGAATG AGCTCGATTA CGGCGACCAG	GCGTGAATGG	TCGGGGAATG	AGCTCGATTA	CGGCGACCAG	14520	
AAACTCGCCT TGCCAGCGCA CCGCAACGTG	CCGCAACGTG	ATCTCCTGGC	TTCACGGTGG	CCCCGACCAC	14580	
AGGTTCTGAC GAGGAATCCG GGGGCCGTCG	GGGGCCGTCG	ລອລລອລລອລອ	AACCACGCGT	ACACCGCCGC	14640	
CACCCAGCCG GTGATCCGGC GGCCGTAGAA AGTGACCGTG GCCACGATGA CGCCCAACGA	GGCCGTAGAA	AGTGACCGTG	GCCACGATGA	CGCCCAACGA	14700	
GGCCAGCGCA ATCCCCGCCC ACCAGTAGCG	ACCAGTAGCG	CGTCTCCAAG	CGTCTCCAAG AATGCGATGA	TGCATGGCGG	14760	
GGCCAACGCG GAGGCAAGCA AGGCGTGCCC	AGGCGTGCCC	GGTGCTGAAC	CGCAGCCCTA	AAGGATTTCT	14820	

						20/6	3						
20 of 22	14880	14940	15000	15060	15120	15180	15240	15300	15360	15420	15480	15540	15600
Page	CCAGGGCCAA CGTAAGGCCG	GACCCGGCTC CACCACCGGG	CGGGCGGAAT GTCCCACGTC	CCGACCAGGT CGTCGACCCC GCCCCCGGGG	ATCTGCGCCG GCGTCAGGTC GGGGAACCGC	TATGCCGCGG CAAACGAGGT GCCGGCGATG	CACCGGTGTC GCCGAGCGCG	CGTGCATCGA GAACGAGCTG	CTTAACACCA GCGGTGCGTA CCACGCCGGG	CGGGTGTGGA CGGGTCCGGC	TTGCCGGCCG CGACCACCAC CACCACGCCT	TTTCATCGAT CGGCCTGCTC	CGCCGAGGTT GGCGGCGTGC
FIGURE 1-20	CATCGGCGGC TCAGCCCCG TCTAGCCAGC GCGCCCAGGC CCAGGGCCCAA CGTAAGGCCG	ACGGCCACCA ACGCCACAGC CGTAATCGGG CGACGATCGG GACCCGGCTC CACCACCGGG	GGTGGAAGTC GTCTGACGTT GTATGGCGCC GAAGCAGGGC CGGGCGGAAT	AGCGCGCCA CCGCATCGAT GACGCCGGCG CCGACCAGGT	TGTCTCGCGG TGGCGGTGAT CCGGTGGATG ATCTGCGCCG	TGCCGAAGCA GGGCCGCCAG ACCCGACACA TATGCCGCGG	GGTACCGGCC CCTCCCGGCC TTGCAACGCA TTCACCGGTT	ACGATGTTTT CTGCGGGCGC GGCCACGTCC ACCCACGGTC CGTGCATCGA GAACGAGCTG	GGCATCCCGG TCTGGCCGAT ACCGCCGACG CTTAACACCA	GTGACAACGG TCTGCACATT GTTCCAGCCG CGTGGGTCGC	GCCGGATTCT GTACGCAATC GCCACCGGTG TTGCCGGCCG	TTGACGTTGA CCGCATAGTC GATGGATGCA CCCAGTGAGG	ACCTTGTAGC AGGCGGCTTC ACTGATGTTG ATCACACCCA

		FIC	FIGURE 1-21		Page 2	21 of 22	
ACCACGGCGC	GGGCAAGACT	ACCACGGCGC GGGCAAGACT GCGGATGGAA CCGGCGGCCG GGGTGGCGTT	ອລລອອລອອລລ		GGGGTCATTC	15660	
GGGTTGGCTT	GTGAGCCGAC	GGGTTGGCTT GTGAGCCGAC CGGTTCGAAG GCCTCAGACG	GCCTCAGACG	TCTGACGTAG CGAGAGCAGT	CGAGAGCAGT	15720	
CGAGCGTCGG	GCGCGACGCC	CGAGCGTCGG GCGCGACGCC GACGAACCCG	TCGGTGGGCG	TCGGTGGGCG CGGCCGGCC CGCGATGATG	CGCGATGATG	15780	
GATGCTGTGA	GAGTCCCATG	GATGCTGTGA GAGTCCCATG GGCATCACAG		TCAGACAGGC CGTTACCGGC CTGGTCGACG	CTGGTCGACG	15840	
AAATCGCCGC	AAATCGCCGC CAGGTTCCGC CGGGACCCGT	CGGGACCCGT	GGCGAAGCGT	CGACACCGGT	GTCGATCACC	15900	
GCCACCGTCA	ລລລລອອລລລລ	GCCACCGTCA CCCCGGCCCC GGTCGCGAAC	TTGTGGGCAT	CGGCCACGCC	CAGATACGTG	15960	21/6
TTGCTCCACG	GCGGATCGTG	GAACCCGGAC	CCCGGCAGCG	TTGCTCCACG GCGGATCGTG GAACCCGGAC CCCGGCAGCG TGGTGGGCGA CGCGCACAAA	CGCGCACAAA	16020	3
ACGCGCTGTT	CGGTAGGCTG	ATCCGGGCCC	GCCACGTCGG	ACGCGCTGTT CGGTAGGCTG ATCCGGGCCC GCCACGTCGG GCGGCAACGC GCCCGGATCG	GCCCGGATCG	16080	
ATCGGCGGTG	ATCGGCGGTG GCGTGATGGC CGATGCGGGC		GACGCGGTGA GCAACGCCAG		CGCCACCGTG	16140	
ATCAGAAAGA	TACGGTGCAC	TCCCAGAACA	CTCCATTCGT	ATCAGAAAGA TACGGTGCAC TCCCAGAACA CTCCATTCGT TGAGATTCAT TGCGATTCAT	TGCGATTCAT	16200	
TGAGCTGCGT	TGCTACCTTG	GGCCACTTGA	CGGACCTGTG	TGAGCTGCGT TGCTACCTTG GGCCACTTGA CGGACCTGTG TGCATTTTAG ACGTAACGGC	ACGTAACGGC	16260	
TGGGCAAACA	TGGGCAAACA ACGCTGTCAC GCCTGGGCTG		SICCECCECE	CCGACCAGGG	CGCGTAGGCG	16320	
CTGTACCTGG	CTGTACCTGG ACCACGCCGG GACTCAACGG		TTTTGCTACC GCACTAGCCG		ATATGCGGCT	16380	

						22/6	3		
22 of 22	16440	16500	16560	16620	16680	16740	16800	16860	16885
FIGURE 1-22	T TGTCTGAGCA CACGCTGCGT ATCGCGGCAT	T CCTGAACCGA TACCGGTTGG CCCGCACGTT	TTTGCGCAAC CACCCGGGTG TCCCGGAACC CTTCGGCGCG TTCGATCACG TTGCGGGCGA	ACCGACCGTT TTGCATAGCG TCGATACCGT GCTGCCCACT AGGGGTGGTG TAGTTACGGA	C GTGCGGCGTC ATCGAGCTGG CTGGCGCGCG	A TCTCCACCGG CGAATAAGAC TCGAACCGCA	GCTTTCGGTT GAACCGGCCA GCCAAACCCG GGTTCACGGT GAGGAATTCG GTACCCCGGG	C TCCCAACGCG TTGGATGCAT AGCTTGAGTA	
EI.	GCTACCAAAC GATCGCGGCC ATGTCTCGGT	CGATGTCGGT GGCGGTGATG ATCTGCAGAT	TTTGCGCAAC CACCCGGGTG TCCCGGAACC	ACCGACCGTT TTGCATAGCG TCGATACCGT	TGGTGGTGAC CGCGTCGAGG AATACCTCCC	GTGTAGCGTA GCGGTGTCCA ATCTCGACGA	GCTTTCGGTT GAACCGGCCA GCCAAACCCG	TTCGAAATCG ATAACTTGGA TCCGGAGAGC	TTCTATAGTG TCACCTAAAT ACTTG

Page 1 of 20	SAGGTC ATCGTCATCG 60	CCCAAG ACACCTGATC 120	CGTGGAGTGT GTTTGGACCA GCAATAGCGT 180	AAGTTATACC CGGTTATACT ATCTGTATGA 240	ATCGCGGCGT GCGAGTGAGC 300	raccig cacgagerge 360	420	ACCATG GACGATGAGT 480	CCGCCATCAG GCAGTCAGCC 540	ACGCAA GTCGCCTTGC 600	PAATGC CCGGCAACGT 660	CAACG TCACGGCGAT 720	いった。日本が担任ができなが、できなどの任じてきない。
FIGURE 2-1	GGATCCTCGG ACTGGCCGCG GTCGTGCTTG TGCACGAGTT CACCGAGGTC ATCGTCATCG	CCAACGGCGT GCGGGCCGGA CGCATCAAAC CACTTGCCGG GCCACCCAAG ACACCTGATC	CGCGCGGAAT CGTGGAGTGT GTTT		TCGATCGGGT	TCGGCATGAG TCGGTCCGAG TTCTTCACGA AGGCTGCGCA GCGCTACCTG CACGAGCTGG	ACGCCCAATT GCTCACGGGC CAGATCGACA GGGCTCTAGA GAGCATCCAT GGCACCGACG	AAGCGGAGGC CCTCGCCGTG GCCAACGCAT ACCGCGTGCT AGAAACCATG GACGATGAGT	TGACCTCGGG	GGCGAAGCGC CGCCCGGTGC TCGTAATCCA GTCAGATCCG TACAACGCAA GTCGCCTTGC	CACTGTGATC GCAGCGGTGA TCACGTCCAA TACGGCGCTG GCGGCAATGC CCGGCAACGT	GITCTIGCCC GCGACCACAA CGCGACTGCC ACGTGACTCG GTCGTCAACG	
	GGATCCTCGG ACTGGCCGCG	CCAACGCCT GCGGGCCGGA	GGACTATCCC GGGGTAGCGA CGCGCGGAAT	CACTGTGACG AAACAGCCGC CGTCTTCTGG	AGACAGCTAT TTCTCTGCCG GATGAGACGT	TCGGCATGAG TCGGTCCGAG	ACGCCCAATT GCTCACGGGC	AAGCGGAGGC CCTCGCCGTG	GGTGATTAGT CGTGCCGAGA TCTACTGGGC	GGCGAAGCGC CGCCCGGTGC	CACTGTGATC GCAGCGGTGA	GTTCTTGCCC GCGACCACAA	

۷	٧	ı	0	J

FIGURE 2-2	Page 2 of 20
GCACGAGGTT GACCGAGGAC TTCGTCGCGT ACTGGACCTT TGACACTGCG CCACGCGACA	ACA 840
ATTCGTCACG GTGACGTTCC TGCTTGGTGT AAGCCCCCCC GCCGGGGGAA CTACTCGCCG	006 5DC
GAGGTGGTGT TGTGGGCAGG CTTGAGGGCA AGGTTGCATT CATTACGGGC GTGGCTCGGG	096 555
GTCAAGGCCG TTCGCATGCG GTCCGCCTAG CCGACGGCCA AGCGCGTGCG CTCGGCAAGG	1020 1020
TCGATGTTGA GGCGTGCGGT GCGCTCGTTG GTGAGGTAGA AGTGTGGGGC CGTGACGTGC	.GC 1080
GTGACGATCG ACGGGTGTTT GTCGAGAGTC CTGCCGACGA GTTCGGCGCG TGCCGCCGCG	tCG 1140
TCGCGCGTCA GGGCATCCGT GTCGTAGGGC TGCCCGTTTC ACAGAGGGAA CTTGTCGAGC	.GC 1200
CCGAAGCCGG GTGCGCGCGCG AGGCGCTCGG CTGCTGGCTC CCAGTAGACA TCTAGGCCTG	TG 1260
CGTCGACTGC GGCTGCGGCA GCGTCGTGCT GGTGACGAGT GGCGTTGGTG TCCAGCGTGA	GA 1320
TCGCAGTGGT GCCGGCGTGG TCGCGGGACA GGAAGTCCTC GACCGGTTTG TGATCACCCG	CG 1380
GCCCGAGCCG AAACTGAATG CCCATCGTCG TGAAGTTCCT CTCGCATCGA CGCCTCGGTT	TT 1440
CGTGTCATAA TACATGACAA ATCAATAGAC AAAAGGAAGA CAGGCTGCCC ATGGGAGTAA	AA 1500
ATGTGCTCGC CTCGACCGTG TCGGGTGCGA TCGAGCGCTT GGGATTGACC TACGAGGAAG	AG 1560

0
ന
Page
, 2-3
FIGURE
FIG

		FI	FIGURE 2-3		Page 3	of 20	
TCGGTGACAT	TCGGTGACAT CGTCGATGCC	TCGCCGCGTT		CCGTGGCGCG ATGGACCGCA GGTCAGGTGG	GGTCAGGTGG	1620	
TTCCCCAACG	TTCCCCAACG CCTCAACAAG	CAACGACTTA	TCGAGCTGGC	CTATGTCGCC	GACGCCCTCG	1680	
CGGAAGTGCT	CGGAAGTGCT GCCGCGTGAC	CAGGCGAACG	TGTGGATGTT	CAGGCGAACG TGTGGATGTT TTCGCCGAAT CGGTTACTGG	CGGTTACTGG	1740	
AACACCGCAA	AACACCGCAA GCCTGCCGAC	CTCGTGCGAG	ACGGCGAGTA	CTCGTGCGAG ACGGCGAGTA CCAACGCGTG TTGGCGCTCA	TTGGCGCTCA	1800	
TCGACGCGAT	TCGACGCGAT GGCGGAGGGA	GTGTTCGTGT	GAGCGATGCC CTCGATGAAG		GGCTCGTCCA	1860	
GCGTATCGAC	GCGTATCGAC GCACGCGGAA	CAATTGAGTG	GTCGGAAACG	TGCTACCGGT	ATACCGGCGC	1920	
GCACCGTGAC	GCACCGTGAC GCCTTGTCCG	GTGAGGGCGC	GCGCAGATTC	GTGAGGGCGC GCGCAGATTC GGAGGCAGGT GGAATCCGCC	GGAATCCGCC	1980	-
GCTGCTCTTT	GCTGCTCTTT CCGGCGATCT	ATCTTGCTGA	TTCCGCCCAA	ATCTTGCTGA TTCCGCCCAA GCCTGCATGG TTGAGGTGGA	TTGAGGTGGA	2040	
ACGGGCGGCG	ACGGGCGCG CAAGCGGCTT	CAACGACCGC	AGAGAAGATG	CTCGAGGCGG	CCTACCGACT	2100	
ACACACGATC	ACACACGATC GACGTCACGG	ACCTGGCCGT	CCTCGATCTG ACAACCCCGC		AAGCTCGGGA	2160	
AGCCGTGGGG CTCGAGAACG	CTCGAGAACG	ACGACATCTA	TGGCGACGAC	ACGACATCTA TGGCGACGAC TGGTCAGGGT GCCAGGCGGT	GCCAGGCGGT	2220	
CGGACATGCG GCCTGGTTCT	GCCTGGTTCT	TGCACATGCA	AGGTGTCCTC	TGCACATGCA AGGTGTCCTC GTGCCGGCGG CGGGCGGTGT	CGGCGGTGT	2280	
CGGCCTCGTT GTCACCGCGT	GTCACCGCGT	ATGAACAGCG	AACTCGGCCG	GGCCAACTAC AACTGCGACA	AACTGCGACA	2340	

						26/6	63						
4 of 20	2400	2460	2520	2580	2640	2700	2760	2820	2880	2940	3000	3060	3120
Page 4	CCACGTAGC TGGCCAGCTT	AGCAACTCG CATCACTTAT	TGGGTTAGC TGCGCCGTTT	CGTGTCGTC AAAGGTTGCG	AATCAGGTAG TTTTAGCCCG	GTCGTACGG CGCAACGACG	GCCGCAATA ATTTGGAACG	TGAGCTACCT CCGGGGTTTC	CATAGAACGA AGTGTGCCAC TTCTAGCAAA	CAGTCACCG CACTTCCGGC	AGGAACGCG CGCCGTGGCA	AGAAGCGAAT TTACATACGA	TTGGGTGCC AGCAGCGTCC
FIGURE 2-4	TCTTTACCA AGAACTTCGA	GTGCCATGG TCATCGTAAG	TCAAGTGAA GCTGGTCGTA	GTGCGTGCT GCGGCAGCGA GCGTGTCGTC	GGCACGCAGC	AGACGCAGC AGCTCGCCAT	TCGAGCATC GATGATCAGG C	GTCCAATGCC	AGTGGTCGA CATAGAACGA A	CGCGGGTTT ACCGCCCCTG C	AATCCACAT GCGGCGGTGG CAGGAACGCG	AACCTTC GCGCGTTGCG	CGGCCGCGA CAAGCCCGCT G
	AAGCGTCGAT CTGACGCCTG CTCTTTACCA AGAACTTCGA GCCACGTAGC TGGCCAGCTT	GGCGCAGAGA AGGATGCCGC TGTGCCATGG TCATCGTAAG GAGCAACTCG CATCACTTAT	AAGCCGATAA GCGACATTAT GTCAAGTGAA GCTGGTCGTA TTGGGTTAGC TGCGCCGTTT	GTGCTAGCGG GCACGCTCCT TGTGCGTGCT	AGGCTTGCCT GGTGATGAAT TGCCACATCC	CTGCTAGCGC GTAGTTCGGC GAGACGCAGC AGCTCGCCAT CGTCGTACGG	CTACCCGCGG AACGCGGATC GTCGAGCATC GATGATCAGG CGCCGCAATA	GGGCTCGCCA GGCCATCGCT GGGCGGCCCG	GGTTATTTGG TAGCGCGGAC GAGTGGTCGA	GGTGGTACAC CTACTGGCGG CCGCGGGTTT ACCGCCCCTG CCAGTCACCG CACTTCCGGC	GGCGGACGAC GAGCACCGCG GAA	GCATGAGGCC GACGGTCAGG GTG	GGGCTCGCTG CAGCACGAGA TCGGCCGCGA CAAGCCCGCT GTTGGGTGCC AGCAGCGTCC

Ď
Pag
14
ŵ
ä
r+1
٢
GURE
-

		FIGURE 2-5		Page 5	Page 5 of 20
GCAGCAGGCG GGACGAAATC	CCGTACGGCG	CCGTACGGCG GGTTCGCCAC AACCCGGAAC GGCCGGCCGG	AACCCGGAAC	ອອນນອອນນອອ	3180
GCAACCGGAT CGAGGCGGCG	TCCGCGTGCA	TCCGCGTGCA CCACGGTAAT GCCAGGGAAT CGCTCGCGGA	GCCAGGGAAT	CGCTCGCGGA	3240
GGACACCGAC TCGTCGCGGG	TGCAACTCCA	TGCAACTCCA CGGCGACCAC CCGCGCCCCC GCTCGCACTA	ວລວລລອລອລລ	GCTCGCACTA	3300
GATGCGCCGT CAGTGCCCCT	TCGCCGGCGC	TCGCCGGCGC CGATGTCAAA CACGAGCTCA CCGGACCGCA	CACGAGCTCA	CCGGACCGCA	3360
CTGCGGCCGC GCTGACTACC	CGCGCTGCCC	CGCGCTGCCC ATTCGTCATG GAGCCGGTGC CAGCCCCATG	GAGCCGGTGC	CAGCCCCATG	3420
CCCGTCGCGA CCGTCCGAGG	GCGGACACGA	GCGGACACGA CGTACCGTCA CTGCGTAGAT GCCCACGCGC	CTGCGTAGAT	GCCCACGCGC	3480
CCGACCGTAG CCCGCCACCG	GCACTGCGAT	GCACTGCGAT CAATCCAATT TCTCGGTTCA GGCAACCTTC	TCTCGGTTCA	GGCAACCTTC	3540
TGGTCATCAC CAGCCCCAGG	GCTCTGGCGC	GCTCTGGCGC CGTCCGCATC AACTCCGAGA TGACGTTGGC	AACTCCGAGA	TGACGTTGGC	3600
CGTGACGACC CACTAGACCC	ACCTGGCAGT	ACCTGGCAGT AGCCGCATTG	TCGCAGTCGG CGAGCCTCAG	CGAGCCTCAG	3660
TGCGCAGTCG CGTCTAGGTG	CAAGGATATT	CAAGGATATT GCCCGTTGAG CAGACAACTC GACGGCGGCG	CAGACAACTC	GACGCCGCCG	3720
AGTAAGAACC GGTCAGCCCG	CCTCTTAGGC	CCTCTTAGGC CGCCCGTGGC TGAACCACCG GGGGCAATGA	TGAACCACCG	GGGGCAATGA	3780
TGCGATTCCA ATTCGCTGGG	CTGAGAACGT	CTGAGAACGT AGTGCGTGCC AGATCGTGCA ACGGTGCTAT	AGATCGTGCA	ACGGTGCTAT	3840
TCCATGTGTG CAAGACGGAT	TCTCCTGCCG	TCTCCTGCCG GCAAGTCGAA TTCAAGCTTC CAATCGGTTA	TTCAAGCTTC	CAATCGGTTA	3900

Page
١.۵
E 2-6
RE
FIGURE

	FIGURE 2-6	Page	e 6 of 20
GCGCCCCT GCTCGAGTTT GTGATGGTC	GTGATGGTGA AGCGGGCGAT GAAACCGGTC TGCCACGTCG	ACCEGIC TGCCACGIC	3960
ATGTCACCGA CAACGTCGCC CTGGCCGTCG	CCGCACTAGC	GACCGGGGTG ATGGCGAGTC	3 4020
CGAGGATGGC AACTATCAAT GCCGACACGG	TTGCGTGAAG	CGCTGTCCGC CAGCGCCTCA	4080
CGTAAATGTT CAGTCCGGCC ATGACAGCC	ATGACAGCCA ACACTAATGC CAATGAGGCG ATATCGGCCG	rgaggcg atatcggcc	3 4140
TCTCCTCGCG AGCAAGCTAC AGCAACTTTG	IG CTCAACCGCA ACCGTGATGA	STGATGA AATTTGGCCT	r 4200
CGACCCACCC TGAACCAGAT ATCGGCCCGG	CCGAACGCGA	ACTTGCGGAC GGGGAAGGCC	3 4260
AGACAGCCTC GACCCCACTC CCCCGATTA	CCCCGATTAG CGCCGTTCAC CGTTCGCGAC CGGTATCAAC	rcgcgac cggtatcaa	3 4320
GGGCTACAGC TCCAACACGA TCCGTAGGGC	CGCGTCACGC	CGAATGTGCA CTGGTGGCGC	3 4380
CGACACGCCC GGGCGAGGCC GCCGTCGGCG	TGTCAGCTGG	TGACTGAGTT GTGCAGACTG	3 4440
ACCGCGCCC CTCCTGCCGA ACGGTATGTG		CCCATCGACG ATCACGTGGT CCAACCCGCG	3 4500
TGTGCACACG TGCTGTACTA GGTCACGGT	GGTCACGGTC AGCGAGATTC CCAG	CCAGCGCAAC CATCATGACC	4560
GCGATCAGGC CGTCGAGGAT TCTCCACGAG	CCGGGGTTGG	TGAACAGCCC GCGCAACCGG	4620
CCGGCTCCGA ACCCGAGGGT GGCGAACCA	GGCGAACCAT ACCGCACTGG CTGTGACCGC GCCGAGGCCG	GACCGC GCCGAGGCC	4680

FIGURE 2-7	Page 7 of 20
AACAGCCAGC GCTGGTCGCT GTGCTCGTTG GCCAGCGCGC CTAGCAACAC GACGGTGTCG	CG 4740
AGGTAGACGT GTGGGTTGAG GAACGTGAAT GCCGCACAGG TCACCAGGAC CTCGGCTAAG	AG 4800
CGAACCGGCG TGGCGCCAGA TGGGATCAGC GCAACAGGTC GCCACGCCCG CCGGGCCGCA	CA 4860
AGTAGCCCGT AGCCGATTAG GAAGGCGGCG CCGCCAAACT TGACGACATT GAGCGCACGC	3GC 4920
GGATGTGCGC CGATCAATGC GCCGAACCCC GCGATACCGG CGGCGATCAG CACGATGTCG	CG 4980
GACACCGTGC ACAGCGCCAC CACCGGCAGC ACGTGCTCAC GCTGGATTCC CTGCCGCAGC	GC 5040
ACGAATGCGT TCTGCGCGCC AATCGCGGCG ATCAGCGTGA AGCAGGCCAG GAAGCCGACG	יכפ 2100
ACCAGTGGTG AGTTCACGCA ATCGACACTA GGCAGTTTGT ATGGGTCAGT ATAGCTAATA	TA 5160
ATTCTTCATT TACATTAGCA TTATTAATGT GCAGTGCGAC GCTCCGCAGA TGGTCTACAC	AC 5220
CTGAGATGGT GGATCCGCAG CTTGACGGTC CACAGCTGGC CGCATTGGCT GCCGTGGTCG	CG 5280
AACTGGGCAG CTTCGATGCG GCCGCGGAGC GCCTACATGT CACCCCCTCG GCTGTCAGTC	TC 5340
AGCGCATCAA GTCGTTGGAG CAGCAGGTCG GCCAGGTGCT GGTGGTCAGG GAAAAGCCAT	AT 5400
GTCGGGCGAC GACCGCAGGT ATCCCGCTGT TGCGGTTGGC CGCGCAAACA GCGTTGCTCG	CG 5460

FIGURE 2-8	Page 8 of 20
AGTCCGAGGC GCTCGCTGAA ATGGGTGGCA ACGCGTCGCT GAAACGCACG	GGATCACCA 5520
TTGCGGTAAA CGCCGATTCC ATGGCGACAT GGTTTTCGGC CGTGTTCGAC	GGTCTCGGCG 5580
ACGTCCTGCT CGACGTTCGG ATCGAGGACC AGGACCATTC CGCGCGGCTG CTACGGGAGG	CTACGGGAGG 5640
GTGTGGCGAT GGGCGCGGTG ACCACCGAGC GGAACCCGGT GCCGGGCTG	GCCGGGCTGC CGGGTGCACC 5700
CGCTGGGTGA AATGCGCTAC CTACCAGTGG CCAGCAGGCC ATTCGTCCAG	GCCATCTAT 5760
CCGACGGGTT CACTGCCGCC GCGCGGCTA AAGCTCCGTC ACTGGCGTGG	AATCGTGACG 5820
ATGGGCTGCA GGACATGTTG GTGCGTAAGG CCTTTCGTCG CGCCATCACC AGACCGACGC	AGACCGACGC 5880
ACTITIGICCC GACCACAGAG GGCTTCACCG CCGCAGCGCG CGCCGGGCTG	GGATGGGGCA 5940
TGTTCCCCGA GAAGCTGGCA GCATCTCCGC TTGCCGATGG ATCGTTCGTA	. CGGGTCTGCG 6000
ACATACACCT CGACGTCCCT CTCTATTGGC AATGCTGGAA ACTGGACAGT	CCGATCATCG 6060
CGCGAATTAC CGACACGGTG AGGGCGGCGG CAAGCGGTCT GTACCGGGGC CAGCAACGCC	CAGCAACGCC 6120
GCCGCCGACC GGGTTGACCG ACGCCAGCAT GTTGTTGTGT CAGCGCGGCT	TGGTCTCGAT 6180
GTCCCGGCCT TGCTGGACCC GCTTCCTCAA ACAGGTTGAA CTTAACGACT	CAGACGGAAA 6240

Page 9 of
6-7
FIGURE 2

FIGURE 2-9	Page 9 of 20
CGCTTGAACC GCGACGTCGC TCCGGACACC AATTTGACTC GGCTCTTTGG CAATTGAAGG	00E9 5500
TGAGCTGCGA GCAGCCGGGT GACCGCATCG TTGGCCTTGC CATCAATCGC CGGCTCGCGG	0989 550
ACGTAGATAA TCAGCTCACC GTTGGGACCG ACCTCGACCA GGGGTCCTTT GTGACTGCCG	3CG 6420
GGCTTGACGC GGACGACCAC AGAGTCGGTC ATCGCCTAAG GCTACCGTTC TGACCTGGGG	3GG 6480
CTGCGTGGGC GCCGACGACG TGAGGCACGT CATGTCTCAG CGGCCCACCG CCACCTCGGT	3GT 6540
CGCCGGCAGT ATGTCAGCAT GTGCAGATGA CTCCACGCAG CCTTGTTCGC ATCGTTGGTG	9TG 6600
TCGTGGTTGC GACGACCTTG GCGCTGGTGA GCGCACCCGC CGGCGGTCGT GCCGCGCATG	ATG 6660
CGGATCCGTG TTCGGACATC GCGGTCGTTT TCGCTCGCGG CACGCATCAG GCTTCTGGTC	3TC 6720
TIGGCGACGT CGGIGAGGCG TICGICGACT CGCITACCIC GCAAGTIGGC GGGCGGICGA	GA 6780
TTGGGGTCTA CGCGGTGAAC TACCCAGCAA GCGACGACTA CCGCGCGAGC GCGTCAAACG	ACG 6840
GTTCCGATGA TGCGAGCGCC CACATCCAGC GCACCGTCGC CAGCTGCCCG AACACCAGGA	GA 6900
TTGTGCTTGG TGGCTATTCG CAGGGTGCGA CGGTCATCGA TTTGTCCACC TCGGCGATGC	0969 ენ:
CGCCCGCGGT GGCAGATCAT GTCGCCGCTG TCGCCCTTTT CGGCGAGCCA TCCAGTGGTT	TT 7020

מע	,	c	7
32	,	ถ	:
~-	•	•	•

10 of 20	7080	7140	7200	7260	7320	7380	7440	7500	7560	7620	7680	7740	7800
Page 1	TATAGCTCTA	AATATTATGG	GCGCGAACA	CTGTAGTCGA	GATGAAATGA	GGGCTGGTGT	GACGGGATGG	CCACCATCGA	GCAGGTAGCG	TGGTTGAGGT	SACCCAACTC	STTGTCGGAT	ATGGACAAGA
	CGGTCCGCTG	CGGAGGCGGC	TCGGGGATGA CAAGCCAGGC GGCGACATTC GCGGCGAACA	TCAAAGACTG TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA	CATCAGGCCG	ATCACCGCCG	GCCGGCAGGG TTCCGGATCC GATGACATAT GACGGGATGG	TCGATGATCG AGGGGACGGG TATGGGAGTC CCACCATCGA	GACGTTGGGC GCAGGTAGCG	GAGCACTCCC	GGAATGCTGG GGCCTGGTCC CACCGCATTG GACCCAACTC	ACGGCACCGA AATAGGCTTC GAACGGGGTT GTTGTCGGAT	GGACGTTGGG 1
FIGURE 2-10	GGCGGGTCGT TGCCGACAAT	CAATATGCAC	CAAGCCAGGC	TTGTCCCTAT	AAGGCAAGA ACCCGGTATT	CGTAGAGCCG	TTCCGGATCC	AGGGGACGGG	CGACGACCTC	GTGCGTCGAT	GGCCTGGTCC	AATAGGCTTC	TGGCGAAGGG
FI	GGCGGGTCGT	CCCGACGATC	TCGGGGATGA	TCAAAGACTG	AAGGGCAAGA	GTGTTGAACG	GCCGGCAGGG	TCGATGATCG	ATCGGCGATC	TCACCTGCGG	GGAATGCTGG	ACGGCACCGA	TAGGTCGAAA
	TCTCCAGCAT GTTGTGGGGC	AGACCATAAA CTTGTGTGCT	CGCATGTTTC GTATGTTCAG	GGCTCGATCA CGCCGGATGA	TGTACACCGG CTGGAATCTG	CGGTCGGGCG GTAATCGTTT	AGACCTCAAT GTTTGTGTTC	TTCCCGTTAC CCCACCGGAA	TCTTTACGTA CAGGGTGGTG	GGTTGGGACC GAACACGAGC	CACCCGGTAA CGCCATCGTC	CCAGAACGCC GTCCACGCCG	CGGCCAGCAA GGCGCTGAAG
	TCTCCAGCAT	AGACCATAAA	CGCATGTTTC	GGCTCGATCA	TGTACACCGG	CGGTCGGGCG	AGACCTCAAT	TTCCCGTTAC	TCTTTACGTA	GGTTGGGACC	CACCCGGTAA	CCAGAACGCC	CGGCCAGCAA

						34/6	3						
2 of 20	8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
FIGURE 2-12	TGCGCGCTCT GCAACTGTGA CGCGATGGTG GCCTCCGCGG CCGCATACGA TCCTGACGCC	GAGTICAGCG TCTGCACGAA CTGTTGATGG AAAGCGCTCG CCTGCGCGCT GACCGCTTGA	TATTCCTGAC CGAACCTGGC AAACAGCGCT GCCACCGCCG CCGATACCTC ATCAGCGCCA	GCGGCCGCAA GCGCGGTGGT CGAGGCGGCA GCCGCGGCAT TCGCCGCGCG CAGTGTGGAA	CCTATGTTCT CCACATCCGC TGCCGCGGAC GTCAAGAACT CGGGAACCAC GACCAGAAAT	GACACGCCGC CCCTCCGCCT CGATCACCAT CCCTGCGCGC ATACAGCGTA TCCAGACGCT	GCCTTTGACA TCTCGGATTT TCAGTAGCTA CCGCACGGCA CAGCACGCGT TAGGTAGATA	GTGGCTATTT GCTGGTACCA TCTACCTGTG GCGCTGAATA TCAAAGACCC TGAGGTAGAC	CGACTAGCCG CCGAACTCGC TGACCGGCTG CACACCAGCA AGACTGCCGC CATCCGGCAT	GCCCTGTCTG CCCAGCTGGC GTTTTTGGAG TCGCGCGCCG GCGACCGTGA GGCACAACTT	CTCGACATCT TGCGTACCGA AATCTGGCCC CTGCTTGCCG ACCGCTCCCC CATCACCAAG	CTCGAGCGCG AACAAATCCT CGGCTACGAC CCCGCAACCG GAGTCTGAGC ACCGCAATGA	TCGTGGACAC AAGCGCCGTG GTGGCCCTGG TTCAAGGCGA GCGGCCGCAC GCCACCCTGG
	TGCGCGGTCT	GAGTTCAGCG	TATTCCTGAC	gcggccgcaa	CCTATGTTCT	SACACGCCGC	SCCTTTGACA	STGGCTATTT	CGACTAGCCG	SCCCTGTCTG	CTCGACATCT	CTCGAGCGCG	rcgtggacac
	TGCGCGGTCT	GAGTTCAGCG	TATTCCTGAC	GCGCCCCCAA	CCTATGTTCT	GACACGCCGC	GCCTTTGACA	GTGGCTATTT	CGACTAGCCG	GCCCTGTCTG	CTCGACATCT	CTCGAGCGCG	

П
Page
д
2-13
RE
FIGURE
F

		FIC	FIGURE 2-13		Page 13	of 20	
TCGCGGCCGC	TCGCGGCCGC CCTGGCCGGC	GCCCATAGCC	GCCCATAGCC CCGTCATGTC	TGCACCCACC GTCGCCGAAT	GTCGCCGAAT	9420	
GCCTGATTGT	GCCTGATTGT CTTGACCGCC	CGTCACGGCC	CCGTTGCGCG	CACGATCTTC GAACGACTTC	GAACGACTTC	9480	
GCAGCGAAAT	GCAGCGAAAT CGGCTTGAGC	GTGTCATCTT	TCACCGCCGA	GIGICATUT TCACCGCCGA GCAIGCCGCI GCCACGCAAC	GCCACGCAAC	9540	
GAGCCTTTCT	GAGCCTTTCT GCGATACGGC	AAGGGGCGCC	Accececeec	AAGGGGCGCC ACCGCGCGC TCTCAACTTC GGAGACTGTA	GGAGACTGTA	0096	
TGACGTACGC	TGACGTACGC GACCGCCCAG	CTGGGCCACC	AACCACTGCT	GGCCGTCGGC	AACGACTTCC	0996	
CGCAAACCGA	CGCAAACCGA CCTTGAGTTC	CGCGGCGTCG	TCGGCTACTG	GCCAGGCGTC	GCGTAACCGT	9720	
ATGCGCGGTG	ATGCGCGGTG ATCGCTGTTT	GTAATGAGTT	CAGCGACACG	GTAATGAGTT CAGCGACACG AAGAATAAAA TATGGGTAGC	TATGGGTAGC	9780	
CGAAATCACT	CGAAATCACT AAGCTACAGT	GCTGGTGCAC	GCCATGAAAG	GCTGGTGCAC GCCATGAAAG ACCGTCAATG ACAAGGAGGA	ACAAGGAGGA	9840	
CGGCCGAAAT	CGGCCGAAAT GCCCAAGGAC	CGACTGCCGG	ACTTGACGCC CACAGGAGCG		TACGCACCGG	0066	
CCAACAGCGG	CCAACAGCGG CATGACCATG	GCAAGGCAGG	ACGCCCTCG	ATGACCGGCA	AGCGCGTTGA	0966	
GCGGGTGCAC	GCGGGTGCAC GCAATCAATT	GGAACCGGTT	GCTCGATGCT	GGAACCGGTT GCTCGATGCT AAAGATTTGC AGGTCTGGGA	AGGTCTGGGA	10020	
ACGTTTGACC	ACGTTTGACC GGTAACTTTT	GGTTGCCGGA	AAAGATTCCG	GGTTGCCGGA AAAGATTCCG CTCTCCAACG ACCTGGCATC	ACCTGCCATC	10080	
TTGGCAAACG	TTGGCAAACG TTGAGTTCCA	CCGAGCAGCA	CCGAGCAGCA GACGACGATC	CGGGTGTTCA	CCGGCTTGAC	10140	

FIGURE 2-14 Page	Page 14 of 20
CCTGCTCGAC ACCGCGCAGG CGACGGTGGG AGCAGTGGCC ATGATCGACG ACGCGGTCAC	10200
CCCCCACGAA GAGGCGGTCC TGACCAACAT GGCGTTCATG GAGTCAGTGC ACGCCAAGAG	10260
CTACAGCTCG ATCTTCTCGA CCCTGTGCTC GACCAAGCAG ATCGACGATG CCTTCGACTG	10320
GTCGGAACAG AACCCTTACC TGCAGCGAAA AGCGCAGATC ATCGTCGACT ACTACCGCGG	10380
TGACGACGCG CTCAAGCGCA AAGCATCGTC GGTAATGCTG GAGTCCTTCC TGTTCTACTC	10440
CGGCTTCTAC CTGCCCATGT ACTGGTCGTC GCGGGGTAAG CTCACCAACA CCGCCGATCT	10500
GATCCGGCTG ATCATCCGAG ATGAAGCCGT CCACGGCTAC TACATCGGCT ACAAATGTCA	10560
ACGAGGTTTG GCCGACCTGA CCGACGCCGA GCGGGCCGAC CACCGCGAAT ACACCTGCGA	10620
GCTGCTGCAC ACGCTCTACG CGAACGAGAT CGACTATGCG CACGACTTGT ACGACGAGTT	10680
GGGCTGGACC GACGACGTTT TGCCCTACAT GCGTTACAAC GCCAACAAGG CGCTAGCCAA	10740
CCTGGGATAC CAGCCTGCAT TCGATCGTGA CACCTGCCAG GTGAACCCGG CCGTGCGCGC	10800
AGCTCTCGAC CCCGGTGCAG GGGAGAACCA CGACTTTTTC TCCGGCTCCG GAAGCTCATA	10860
CGTAATGGGC ACCCACCAAC CCACCACGA CACCGACTGG GACTTCTAAC CGCCCAGCGC	10920

15
E 2-1
$\overline{}$
CRE

	FIC	FIGURE 2-15		Page	15 of 20	
SAGCACCA	CGCGACACCG	GGCCCGATCG	GTCGGGGGCG TCGAGCACCA CGCGACACCG GGCCCGATCG ATCTGCTAGC TTGAGTCTGG	TTGAGTCTGG	10980	
GTCAGCAG	TCAGGCATCG TCGTCAGCAG CGCGATGCCC		TATGTTTGTC GTCGACTCAG	ATATCGCGGC	11040	
CGCCTGCG	AATCCAATCT CCCGCCTGCG GCCGGCGGTG	CTGCAAACTA CTCCCGGAGG		AATTTCGACG	11100	
TCTTCATG	TGCGCATCAA GATCTTCATG CTGGTCACGG	CTGTCGTTTT	CTGTCGTTTT GCTCTGTTGT TCGGGTGTGG	TCGGGTGTGG	11160	
CCAAGACC	TACTGCGAGG	AGTTGAAAGG	CCACGGCCGC GCCCAAGACC TACTGCGAGG AGTTGAAAGG CACCGATACC GGCCAGGCGT	GGCCAGGCGT	11220	
TGTCCGAC	CCGGCCTACA	GCCAGATTCA AATGTCCGAC CCGGCCTACA ACATCAACAT	CAGCCTGCCC AGTTACTACC	AGTTACTACC	11280	37/ 63
CGCTGGAA	CCGACCAGAA GTCGCTGGAA AATTACATCG	CCCAGACGCG	CGACAAGTTC	CTCAGCGCGG	11340	
CTCCACGC	GAAGCCCCCT	ACGAATTGAA	CCACATCGTC CACTCCACGC GAAGCCCCCT ACGAATTGAA TATCACCTCG GCCACATACC	GCCACATACC	11400	
CGCCGCGT	AGTCCGCGAT ACCGCCGCGT GGTACGCAGG	CCGTGGTGCT	CCGTGGTGCT CAAGGTCTAC CAGAACGCCG	CAGAACGCCG	11460	
CAACGACC	GCGGCACGCA CCCAACGACC ACGTACAAGG	CCTTCGATTG	GGACCAGGCC	TATCGCAAGC	11520	
ACACGCTG	TGGCAGGCTG	ACACCGATCC	CAATCACCTA TGACACGCTG TGGCAGGCTG ACACCGATCC GCTGCCAGTC GTCTTCCCCA	GTCTTCCCCA	11580	
AACTGAGC	AAGCAGACCG	TTGTGCAAGG TGAACTGAGC AAGCAGACCG GACAACAGGT	ATCGATAGCG CCGAATGCCG	CCGAATGCCG	11640	
TGAATTAT	GCTTGGACCC GGTGAATTAT CAGAACTTCG	CAGTCACGAA CGACGGGGTG		ATTTTCTTCT	11700	

						J 0 / U .	,						
16 of 20	11760	11820	11880	11940	12000	12060	12120	12180	12240	12300	12360	12420	12480
Page	CCGGCCCAAC CCAGGTATTG GTCCCACGTT	GCCTAGACTC GCGAGGACCG CGCGGTGGTC ACTGCGCGGA	TGTTCGGTGC GCCCACTGCG GTGACTCACC TGCAGCGCCG	TCTATGGTGC GTTAGAGGAT	CGAGAGGATA TGCGATCCAC	CGTCATGCTC GGGATCAACT CGATAATCGG CGCCGGTATC	TCGCGCCGAT GGCCTATGTT	CGACGCCGC AAGGTACGTC	ACGGCCGCAT TTGGGCGCCG GATCGGCATC	GGGGGGTGTT GGCTTCTTTT	CCTGGGCCGA CGCCGAGCAA	GCGTGCTGTT GGCCATCAAC	CAAGTGGGCC AACGGAACGT CAACGGTAGG CAAGGCATTC
FIGURE 2-16	TCAACCCGGG GGAGTTGCTG CCCGAAGCAG CCGGCCCAAC	CCGCGATCGA CTCGATGCTG GCCTAGACTC GCGAGGACCG	TTTGGGGCGG CGGAAGTGAG TGTTCGGTGC GCCCACTGCG	GCATCGACAG GCCGGGAGCT CAAGAATCGT CGCTAGAGAA	TCCCTGCTAG ACAGCCTTGG TGCGGTGGTC GGCCCGCGGA	AAGCTGGGTT TCTGCAGCGT CGTCATGCTC GGGATCAACT	TTCCTAACTC CAGGTGAGGT GATCGGGCTC GCAGGACCCT	TTAGCTGGCA TTTTCGCGGG TGTCGTGGCG ATCGTCTTCG	AGAACAAACG GTGCCTCCTA CGCCTACACA ACGGCCGCAT	TATGTCGGTG TCACCCACGC CATTACCGCG TCCATCGCTT	TTCGTCTCGA CGCTGTTGCG AGTGGCCTTC CCCGACAAGG	CTGTTCAGTG TGAAGACGCT GACGTTTCTC GGCTTTATCG	CTCTTCGGCA ACCGGGCGAT CAAGTGGGCC AACGGAACGT

						39/6	3						
17 of 20	12540	12600	12660	12720	12780	12840	12900	12960	13020	13080	13140	13200	13260
Page	CGTGAACAAC	CGTCGCCGAA	GTACGCATTC	CCGGAACCTG	CACCCTAACG	GAAACTGGCC	GATATCGATG	GTTAGCGGAC	GATGGTCTCC	CGACAACCTG	GCCGATCGCT	AAATGCGTTC	AGTGTCCTAC
	TGGATCATCA CCACCCAGCA	CGTTGCTTGG	ATTGGCAAGG GCACGTTCTC GAGTATGGCG CTGGCCACGA TTGTCGCGTT GTACGCATTC	ACCGGTTTCG AATCGATCGC GAACGCCGCC GAAGAAATGG ACGCGCCGGA CCGGAACCTG	TCTACTTGCT	GCGACACCGT	GCGGCCATCG GAAACGCTAC CTTCCGAACG ATCATCGTCG TCGGAGCCCT GATATCGATG	GCACCGCGGC TTTGGACCGC GTTAGCGGAC	ACGACGTGCC	CGCTGCGGTT	CACCTGACCG GCCTGGCGGT GATCGCCCGA TTCGTCCAGT TCATCATCGT GCCGATCGCT	CTGTGCGGCG AAATGCGTTC	TTGGGCTGGC
FIGURE 2-17		ACCCCGTACT	CTGGCCACGA	GAAGAAATGG	GTTGGCGCGA	GCCGCGTCGG	ATCATCGTCG		AAGAACCAAT	TTCCCGTTGG	TTCGTCCAGT	GAACATGCTG	GTGGTCTCGG
FI	CGGCGGGCTG	ATACAGCGCG	GAGTATGGCG	GAACGCCGCC	GATCTTCTCG	GAACAAGATC	CTTCCGAACG	CTCGTTCGGT	CTTGTCACGC	GGCGCTCGCA	GATCGCCCGA	TCAGGCAGTA	TGTTGCGATC
	GCGCTCTCGG CATTCATTGT CGGCGGGCTG	TACGCAACGG CGTGGTCGGC ATACAGCGCG	GCACGTTCTC	AATCGATCGC	CCGAGAGCTA TACCGATCGC GATCTTCTCG	GTAGCGATGC TGCTCGGATC GAACAAGATC	GAAACGCTAC	Tregeratea atgregege crestreger	AGCGGGGTTC TGCCGACACG CTTGTCACGC	TTCGCAATTA CGGCGTCGTT GGCGCTCGCA TTCCCGTTGG	GCCTGGCGGT	CTCATCGCAT TGGCGAGGTC TCAGGCAGTA	ACCGACAAGG TGTTACCGCT TGTTGCGATC
	GCGCTCTCGG	TACGCAACGG	ATTGGCAAGG	ACCGGTTTCG	CCGAGAGCTA	GTAGCGATGC	GCGGCCATCG	TTCGGCATCA	AGCGGGGTTC	TTCGCAATTA	CACCTGACCG	CTCATCGCAT	ACCGACAAGG

40/	63
-----	----

FIGURE 2-18	Page 1	18 of 20
GACTACCGCT GCATCTTTCT AGTGCGGGGT GGTCCGAACT ACTTCT	ACTTCTCGAT TGCTTTGATC	13320
GTGATCACGT TCATCGTGGT ACCGGCGATG GCTTATCTGC ACTACTACCG AATCATTCGC	ACCG AATCATTCGC	13380
CGGGTTGGCG ATCGGCCGAG CACTCGCTAG ATTCCGTTGG CGCTGAGCTC GAACGGGAGA	CTC GAACGGGAGA	13440
ACACAACGGC GAGCGATGGC GGGAATAGCC TGGTCGGTGC GGGCAAGATT	SATT TCAACCTGCA	13500
TTCCCGGATC GGCGGCGCGG GCAAGCGTCT GCAACGCCGA GGGACTGTAG	TAG GCACGTAGTG	13560
CGCTGATAAA GCCGTCGTGC ATGCTCGAGC GCATCGACGA CCATGGCAGC AGCAGTAGGT	AGC AGCAGTAGGT	13620
GGAGCGGCAG TAGCAGCACC GAAGAGAGCG TGAACGACAG CGGTTTCTGC CGTTTGAGGT	TGC CGTTTGAGGT	13680
CGATGATCAG AAAGCGCTTC CCCACCCGGG TGGCCTCGGC GATCGCTTTG	TTG CAGGCGACCG	13740
TAGGCGGCAG GTGGTGAAAT GCCAGCGCGA AGACCGCCAG GTCATAGCTG	CTG TGGTCGTGGC	13800
CGTCGATTGC GGTGGCGTCG ATCACTTGGG TGCGTGCTCG CGGATGTGTT	GTT CCCAGCTCTC	13860
CCGCGGCGAT GTTGGCCACC GAGGTGGGAT CTAGATCGCT GATCGTCACC	ACC GTCGCTGTCG	13920
GGTGTAGCTC GAGGATTTTC GCTGAGAGCT TGCCATGGCC CGCACCAAGT	AGT TCCAGGATTC	13980
GCGGGTTGGG AATGTCAGAA ACAAGTTTCA GGGCTATCCG GGCGTACTTC TCGTGCAGGT	TTC TCGTGCAGGT	14040

		FIC	FIGURE 2-19		Page .	Page 19 of 20
TGGTCAGGGT	TGGTCAGGGT GCCCACCCGG	TCGAGCACCC	TCGAGCACCC CGATGATCTT	CTGTTTGACC TCATCGGGCA	TCATCGGGCA	14100
CATCGTCGCG	CATCGTCGCG GTCGAGGTAC	TCCAGTGCGT	CGGTCTGGAA	TCGACGATCC	AGCCAAGACG	14160
CGTCGGGGCC	CGTCGGGGCC ACCCCGTGGC	ATCGTGGCGA	TCGCCTGCTC	GCGGATGTTC	GCCTCACCCA	14220
TGGCAGCTCT	TGGCAGCTCT TCCCCTCTCG	ACGTCCCGTG	TTCGCAATGC	TTCGCAATGC TATGAGACCG CTGACCGGGC	CTGACCGGGC	14280
TCCCCAGCCC	TCCCCAGCCC GCCGGTCGCG	CGTGCTTAGC	TACGTAGCAG	CGTGCTTAGC TACGTAGCAG AGGGGCCGTC ACTTCGAGGG	ACTTCGAGGG	14340
CTGCCGCCAC	CTGCCGCCAC TCGGTGATCT	TGCGGCCCAA	TGAATCGGCC GCGTTCGAGG		CTGCCCGTCC	14400
CACGGCTTTG	CACGGCTTTG GTTCACGGTG	AAGATCGCAC	AGCCGGTGCC	AAGATCGCAC AGCCGGTGCC GGAAAAGTCC GCGGCACCGA	GCGGCACCGA	14460
rgtcggtcag	IGTCGGTCAG CAAGACGTTG	AAGAGAAACC	CCGAGATCAC	AAGAGAAACC CCGAGATCAC CGCCCATGGG ATCGTCATCA	ATCGTCATCA	14520
ACACCCCAGG	ACACCCCAGG CAGCGTCGAC	ACCCGCGCCA	ACCCGCGCCA CGAACCAGCA CTGAAGTAGG		TATTCACGCC	14580
ACGCGAAAGG	ACGCGAAAGG CGGCTTGAAC	ATGCACACGG	ACGTGTCGAG	CGTCATCGCG	AAGAAATCGC	14640
CCAGCGCGCC	CCAGCGCGC CACCGGCCGC	AAGACCGGAT	CAGCGACCCG	AAGACCGGAT CAGCGACCCG ACCGGCCGCC TTGTCGGCCA	TTGTCGGCCA	14700
CGATTACCAT	CGATTACCAT GGCGCGCGC	ACCAGCTGGA	TTCGATGCTG	ACCAGCTGGA TTCGATGCTG GGCCGTTGGG TGAGGTGGCG	TGAGGTGGCG	14760
CACGCTGGCC	CACGCTGGCC CCCCGGACAG	GTCGACGATC GGTGACATTG	GGTGACATTG	GTGAGCGTAC GCGGCAGAGA	GCGCAGAGA	14820

CAATGC 14880	GATGTG 14940	TCACGC 15000	TGCCGC 15060	TTGCCA 15120	ACTCGC 15180	AATTC 15239
TGATCGGTAA ATAC	CTATAGCGCG GGGC	GCCGTTGGCT CGAC	GGGCGGTGAT CGG	CGCGACCGGC GAGC	TTTTGCCAGC CCCCACTCGC	CGTTGCGCAC GGCG
TACGCGATTG CTTGGACAAC TGATCGGTAA ATAGCAATGC	TGATGTATCT TGCTAGTATC CTATAGCGCG GGGCGATGTG	ACAGGCGCAT CACCGGTCAA GCCGTTGGCT CGAGTCACGC	GCATCAACAG CGCGCCCGAC GGGCGGTGAT CGGATGCCGC	GGCCCGCCGA CCAGAGCCTT CGCGACCGGC GAGGTTGCCA	TGCTGCTAAC GAGCCTGTAG	GGCTCAGCGA CGGCTCATGT CGTTGCGCAC GGCGAATTC
CCGCTGATGT CCATAGCCAA	AAACTGGCAT ATATTGGCTA	CTCTGCTGCC TTGGCGGCCG	TGGCGAGGCA CCACGATCAG	ATCCTGACCG CCTCGATTCG	CCATGGTCGT CGAAGCAACT	GCTTTGTCTG CAGGTTTTCA

						43/6	3						
1 of 16	09	120	180	240	300	360	420	480	540	009	099	720	780
Page	GAATTCACTT AGCTAACACC AGTTCTAGCA GCTGTCGGCG CGACTTCTTG TCAGTGCCCG	TCTTTGATGC	GCTGGATGCC GAGCTGGACC GCTTGGACGA GGTGTCTTTT GAGGTGTTGA CCACCCCGGA	CTGGAATG CTTGGTGCGC CGGCTACCGG CGGTCGGGCA	CAGCGAGGAA GAACTGGGCG GCACGCTGTG	CAAGCCCGAC GCCGCCCTAC GCATCGCCGA	CGCCGCCGAT CTCGGACCTC GTCCGAGCAC TCACCGGCGA ACCGCTAGCC CCACAGTTTG	TCAAAGTGAT	GTGTCCAAAC CCGCCAGGCC	GGCCCGCTAC	CGAACGCGCC	GCTAAGTGGC	TAGCCAAACT GGCCGCCCCC
	CGACTTCTTG	CTTGTCGAGG	GAGGTGTTGA	CGGCTACCGG	GAACTGGGCG	GCCGCCCTAC	ACCGCTAGCC	AAGGCGCACA	GTGTCCAAAC	CCGACGAGCT	TCACCGACAC	GCATGTCACG	
FIGURE 3-1	GCTGTCGGCG	CCGGGAGGAG	GGTGTCTTTT	CTTGGTGCGC	CAGCGAGGAA	CAAGCCCGAC	TCACCGGCGA	CTGATCGGCG	CGCGGTGGAT	AAATATCGTC	GACGGCGACC	CAATACGACG	GAAGCCGTGC
FI	AGTTCTAGCA	TAGCGAATAG	GCTTGGACGA	GTCTGGAATG	ACACCCAAGC	TACGCATCAC	GTCCGAGCAC	ACGCCAGGGC	ACCTGCCCGC	CAAACCGCTC	GCTACACCCC	GAGCAACCAG	
	AGCTAACACC	ACGTTATGAT TCGAACATGT TAGCGAATAG	GAGCTGGACC	ACGGCTGCGG TCTCTGGAAC GT	CACGTTGATC AACCAACTCG ACACCCAAGC	CTGCGCGCTG GCCAACCGGT TACGCATCAC	CTCGGACCTC	ACCGCCACCG CCACCGCCCA ACGCCAGGGC CTGATCGGCG AAGGCGCACA TCAAAGTGAT	TCGCGCCCTT TTTCGGCCCA ACCTGCCCGC	GCCGAAGCCC GACCTGGCCG CAAACCGCTC AAATATCGTC CCGACGAGCT	GCCCAGCGGG TCATGGACTG GCTACACCCC GACGGCGACC TCACCGACAC CGAACGCGCC	CGCAAACGCG GCATCACCCT GAGCAACCAG CAATACGACG GCATGTCACG GCTAAGTGGC	TACCTGACCC CCCAAGCGCG GGCCACCTTT
	GAATTCACTT	ACGTTATGAT	GCTGGATGCC	ACGGCTGCGG	CACGTTGATC	CTGCGCGCTG	CGCCGCCGAT	ACCGCCACCG	TCGCGCCCTT	GCCGAAGCCC	GCCCAGCGGG	CGCAAACGCG	TACCTGACCC

FIGURE 3-2	e 2 of 16
GGCGCGACCA ACCCCGACGA CCACACCCCG GTCATCGACA CCACCCCCGA TGCGGCCGCC	3 840
ATCGACCGCG ACACCCGCAG CCAAGCCCAA CGCAACCACG ACGGGCTGCT GGCCGGGCTG	006
CGCGCGCTGA TCGCCTCCGG GGAACTGGGC CAACACAACG GTCTTCCCGT CTCGATCGTG	096
GTCACCACCA CCCTGACCGA CCTGCAAACC GGCGCCGGCA AGGGCTTCAC CGGCGGCGGC	1020
ACCCTGCTAC CCATGGCCGA TGTGATCCGC ATGACCAGCC ACGCCCACCA CTACTCCCCC	1080
GCAAGCGGGA GGTACCCCCA GGCGATCTTC GACCACGGCA CACCCCTGGC GCTGTATCAC	1140
ACCAAACGCC TAGCCTCCCC GGCCCAGCGG ATCATGCTGT TCGCCAACGA CCGCGGCTGC	1200
ACCAAACCCG GCTGTGACGC ACCGGCCTAC CACAGCCAAG CCCACCACGT CACCGGCTGG	1260
ACCAGCACCG GACGCACCGA CATCACCGAC CTCACCCTGG CCTGCGACCC CGACAACCGA	1320
CTCGCCGAAA AAGGCTGGAC CACCCGCAAA AACACCCACG GCCACACCGA ATGGCTACCA	1380
CCACCCCACC TCGACCACGG CCAACCGTGG ACCTGTGAGA TACACTACAC	1440
TGCTGTCTAC CTCCGAATCT CAGAAGACCG CTCCGGCGAA CAGCTCGGCG TGGCCCGCCA	1500
ACGCGAGGAC TGCCTAAAGC TGTGCGGGCA GCGAAAATGG GTGCCCGTCG AGTACCTCGA	1560

3-3
JRE 3
FIGUI
E

	raye 3 OL 10
CAACGACGTC AGCGCATCAA CCGGCAAGCG CCGCCCCGCC	1620
CATCACCGCC GGCAAGATCG CCGCCGTGGT GGCCTGGGAC CTGGACCGGC TCCATCGCCG	1680
TCCCATCGAG CTGGAAGCCT TCATGTCATT AGCCGACGAG AAGCGGCTGG CCCTGGCCAC	1740
CGTCGCCGGC GACGTTGACC TGGCGACACC CCAGGGCCGG CTAGTCGCCC GCCTGAAGGG	1800
GTCGGTGGCC GCTCACGAAA CCGAGCACAA GAAGGCACGA CAGCGCCGCG CCGCCCGC	1860
GAAAGCTGAA CGCGGCCACC CCAACTGGTC GAAAGCCTTC GGCTACCTGC CCGGCCCCAA	1920
CGGTCCCGAA CCCGACCCCC GGACAGCGCC GCTGGTCAAA CAGGCCTACG CCGACATCCT	1980
CGCCGGGGCG TCCCTGGGCG ACGTGTGCCG CCAGTGGAAC GACGCCGGGG CGTTCACCAT	2040
CACCGGCCGC CCGTGGACGA CTACAACGCT GTCGAAATTC TTGCGCAAAC CCCGCAACGC	2100
CGGACTACGC GCATATAAGG GTGCCCGCTA CGGCCCGGTG GACCGCGACG CGATTGTCGG	2160
CAAGGCCCAG TGGTCGCCGC TGGTGGACGA GGCGACGTTC TGGGCCGCCC AGGCCGTGCT	2220
GGACGCCCCC GGCCGCCCC CCGGCCGCAA AAGCGTGCGC CGCCACCTGC TGACCGGGCT	2280
GGCAGGCTGC GGCAAATGCG GCAACCACCT GGCCGGCAGC TACCGGCACCG ACGGCCAGGT	2340

Page
Ä
₽
3
H
K
FIGURI
H

CGTCTACGTG TGCAAGGCG	TGCAAGGCGT	T GCCACGGGGT GGCCATCCTG GCCGACAACA TCGAACCGAT	GGCCATCCTG	GCCGACAACA	TCGAACCGAT	2400
CCTGTATCAC ATCGTGGCCG	ATCGTGGCCG	AGCGGCTGGC		CATGCCCGAC GCCGTTGACT	TGTTGCGCCG	2460
GGAGATTCAC GACGCCGCCG	GACGCCGCCG	AAGCCGAAAC	CATCCGCCTG	GAACTGGAAA	CCCTCTACGG	2520
GAGCTGGACA GGCTCGCCGT	GGCTCGCCGT	CGAACGCGCC	CGAACGCGCC GAAGGGCTAC	TGACCGCGCG	CCAGGTGAAG	2580
ATCAGCACCG ACATCGTCAA	ACATCGTCAA	CGCCAAGATA	ACGAAACTTC	CGCCAAGATA ACGAAACTTC AGGCCCGCCA	ACAGGATCAG	2640
GAACGGCTCC GAGTGTTCGA	GAGTGTTCGA	CGGGATACCG	TTGGGAACAC	CGGGATACCG TTGGGAACAC CGCAAGTCGC CGGGATGATA	CGGGATGATA	2700
GCCGAGCTGT CGCCGGACCG	CGCCGGACCG	GTTCCGCGCC	GTCCTCGACG	GTTCCGCGCC GTCCTCGACG TCCTCGCTGA AGTCGTTGTC	AGTCGTTGTC	2760
CAGCCGGTCG GCAAGAGCGG	GCAAGAGCGG	CAGGATATTC	AATCCCGAAC	CAGGATATTC AATCCCGAAC GGGTGCAGGT GAATTGGCGA	GAATTGGCGA	2820
TGAGCCGGCA CCACAACATC	CCACAACATC	GTGATCGTCT	GTGACCACGG	GTGATCGTCT GTGACCACGG CCGCAAAGGC GATGGCCGCA	GATGGCCGCA	2880
TCGAACACGA GCGCTGCGAT	GCGCTGCGAT	CTTGTCGCGC	CGATCATTTG	CTTGTCGCGC CGATCATTTG GGTCGACGAG ACCCAGGGCT	ACCCAGGGCT	2940
GGTTACCGCA GGCGCCAGCG	GGCGCCAGCG	GTGGCAACAT	TACTCGACGA	TACTCGACGA CGACAACCAG CCGCGAGCCG	CCGCGAGCCG	3000
TTATTGGCTT GCCGCCCAAC		GAGTCTCGCC	TACGACCTGA	AATGCGCCGC	GACGGGTGGG	3060
TGCGGCTGCA CTGGGAATTC		GCCTGCCTGA	GCCTGCCTGA GGTACGGCGC CGCCGGCGTG		CGCACGTGCG	3120

Ċ
3-5
4
r-1
-
\simeq
GURE
9

FIGURE 3-5	Page 5 c	of 16
AGCAGCGCC CGTGCGGGTT CGCAACGGCG ACCTGCAAAC	ACCTGCAAAC ACTGTGCGAG AACGTTCCGC	3180
GGCTACTGAC CGGACTGGCC GGCAACCCCG ACTACGCACC	GGCAACCCCG ACTACGCACC GGGTTTTGCG GTGCAGTCGG	3240
ACGCGGTGGT CGTCGCCATG TGGCTGTGGC GCACGCTCTG	CGAAAGCGAC ACGCCGAACA	3300
AACTACGCGC CACCCCAACG CGTGGTAGCT GCTAGACTCC	GACGTAGCCG GCTTCGACTC	3360
CGGGGTTTTG GTGTCCCCAA GGAGTCGCAC GTGTCGACCA	GGAGTCGCAC GTGTCGACCA TCTACCATCA TCGCGGCCGC	3420
GTAGCCGCAC TGTCTCGTTC CCGCGCATCC GACGATCCCG	AGTTCATCGC CGCGAAAACC	3480
GATCTCGTTG CCGCGACCAT CGCGGACTAC CTCATCCGCA	CCCTCGCCGC AGCGCCGCCC	3540
CTGACTGACG AGCAGCGCAC CCGGCTGGCC GAGCTGCTGC GCCCCGTGCG GCGGTCAGGC		3600
GGTGCCCGAT GACCGCCGGC GCCGGCGGGT CGCCCGCCGAC GCGACGATGC	TCGGCCACGG	3660
AGGACCGGGC ACCCGCGACA GTCGCCACAC CGTCTAGCGC CGATCCTACC	GCGTCACGCG	3720
CCGTGTCGTG GTGGTCGGTG CACGAGCATG TCGCGCCGGT	TCGCGCCGGT CCTGGATGCT GCCGGGTCGT	3780
GGCCGATGGC CGGCACACCG GCCTGGCGTC AGCTCGACGA CGCCGATCCT	CGCAAATGGG	3840
CCGCGATCTG CGACGCAGCC CGGCACTGGG CTCTGAGGGT AGAGACGTGC	CAGGAGGCGA	3900

TGGCGCAGGC GTCACGTGAC GTATCTGCGG CCGCCGACTG GCCCGCATC GCCCGCGAGA 3960 TCGTCCGACG GCCGGGCGT TACATCCCGG GCCGGGGGT GGCGTGATGG CCGACATCCC 4020 CTACGGCACC GACTATCCCG AGCCCCCTG GATCGACGG GACGGGCACG TGCTCATCGA 4080 CGACGGTGGC AAACCGACG AAGTTCATCG CGGCCAAGCC GAATCGCCT ACCGGCTAGC 4140 CGAACGTTAC CAGGACAAGC TGCTGCACGT GGCCGAGACT CCTGGGACGG 4200 CGAACGTTAC CAGGACAAGC TGCTGCACGT GGCCGAGATC GCAGTGCTGG AAATGCGG AAACGGTCC GAAACGTTAC CAGGACGTCG GCTGCCGACGCC GCCGACGCTCC GAAATTACGC GCCGACGCTCC GAAATTACGC GCCGACGCTC GCCGACGCTC GAAATTGCCG GCCTGCCCG GCTGCTCGA GCCTGCCCC GCCTGGCTC GAAATTGCCG GCCTGCCCC GCCTGCTCC GAAATTGCCG GCCTGCTCC GAAATTGCCG GCCTGCTCAACGCTC GCCTCGCACC GCCTGCTCCC GCCTTCGCCCC GCCTTCGCCCC GCCTTCGCCGC GCCTTCGCCCC GCCTTCGCCCC GCCTTCGCCCC GCCTTCGCCCC GCCTTCGCCCCC GCCTTCGCCCCC GCCTTCGCCCCC GCCTTCGCCCCC GCCTTCGCCCCCCCC
at at the state of
TGGCGCAGGC GTCACGTGAC GTATCTGCGG CCGCCGACTG TCGTCCGACG GCGCGGCGTG TACATCCCGC GGGCGGGGGT CTACGGCCAC GACTATCCCG ACGCCCCCTG GATCGACCGG CGACGGTGGC AAGTTCATCG CGGCCGAGCC CGACGGTGGC AAGTTCATCG CGGCCAAGCC CGACGGTGGC AAGTTCATCG CGGCCAAGCC CGACGGTGGC AAGTTCATCG GGCCGAAGCC CGACGGTGGC AAGTTCATCG CGGCCAAGCC CGACGGTGC AAGTTCATCG CGCCCAAGCC CGACGGTGC AAGTTCATCG CGCCCAAGCC CCAAGCGTTAC CAGGACCGACG ACGCCGAAACGT CCAAGCGTTC GGCGTGGCCG GCGTGCTCGA GGAATTACGC GCCTGCACGC GCCTCGACA GCGCCCGCTGAC CCTGCACACG CTCAAATTGC GGCCCCACGC CTTGCTCAAC CCTGCACACG CTCAAATTGC GGCCCCACGC TCTCTGGCAA GCCCGATGAA GGTGTGCGCG GCTTCGTGCA GCCCGATGAA GGTGTGCCG GCTTCGTGCA GCCCGATGAA GGTGTGCCG GCTTCGTGCCAA GCCCCGATGAA GGTGTGCCG GCTTCGTGCCAA

		Ī	FIGURE 3-7		Page 7	of 16	
CGACAAGGCG	CGACAAGGCG ATTCGCTATG CC	CCCTTGGCGA	TTATGCCTGC	TTATGCCTGC ACCGCTGAGC CTGACCTTTT	CTGACCTTTT	4740	
CATGCACCGG	CATGCACCGG GAAAACGCTC AC	ACCCAACAGG	CGAAATGGAC	CCAACAGG CGAAATGGAC CTCCGCGGCG TGCGATGGGT	TGCGATGGGT	4800	
AGCGGTATCC	AGCGGTATCC GAGAGCGAAA AA	AAGATCGCCG	GCTGGCCGAA	GATCGCCG GCTGGCCGAA TCAACGATAA AACGGCTGAC	AACGGCTGAC	4860	
TGGCGGCGAC	TGGCGGCGAC GCCATCCGCG CC	CCCGAAAGAT	GCGGCAAGAC	TTCGTGGAAT	TCGAGTGGTG	4920	
CCGTTTGAAG	CCGTTTGAAG TAGTGATTCC	TGCCGACGAG	CAGGACCGGG	AACTGGACGC	ACGGTTGCAG	4980	
TTGGAGGCCG	TTGGAGGCCG ACAGCATCCT GT	GTCCTGGGCG	GTGGCCGGAT	CCTGGGCG GTGGCCGAT GGAGCGACTA TCAGCGAATC	TCAGCGAATC	5040	49/
GGACTATCCC	AGCCGGACGC	GGTGCTCGCG	GCAACGTCGA	GGACTATCCC AGCCGGACGC GGTGCTCGCG GCAACGTCGA ATTACCGCGA GGACTCCGAC	GGACTCCGAC	5100	63
ACGATAAAGA	ACGATAAAGA GGTTCATCGA CGACGAATGC	CGACGAATGC	GTCACCAGCT	CGCCGGTGCT GAAAGCCACT	GAAAGCCACT	5160	
ACTACGCATC	ACTACGCATC TGTTCGAGGC GTGGCAAAGG	GTGGCAAAGG	TGGCGGGTGC	AAGAAGGCGT	ACCCGAAATC	5220	
TCGCGCAAAG	CGTTCGGCCA	GTCGCTCGAC	ACCCACGGAT	TCGCGCAAAG CGTTCGGCCA GTCGCTCGAC ACCCACGGAT ACCCGGTCAC TGACAAGGCC	TGACAAGGCC	5280	
CGTGATGGTC	CGTGATGGTC GTTGGCGGGC CGGAATAGCG		GTGAGAGGGG	GTGAGAGGG CCGATGATTT CGATGATTAG	CGATGATTAG	5340	
CACACCTAAC	CACACCTAAC GTGACGCATG TGACGCATTT	TGACGCATTT	CCAGGTTCGC	CTACGCGCGC	GCACGTATGG	5400	
CGGTTATACC	CGGTTATACC GCGCAAACGT	CACATGCGTC	ACGGCCTGCC GTGCCGTTCT		GCCCAGGATG	5460	

						50/6	3						
8 of 16	5520	5580	5640	5700	5760	5820	5880	5940	0009	0909	6120	6180	6240
Page	CGGTACCTAC CTGGCCGTTC ACGGCCGCCA CCGGGCGGAC TGTACCGCCA AACCAGCAAA	TCGGCACAGC	TCGTTGATCA AGTCCTCGAT CCCCGCCTAT	CCCGCTGC TGGTTCATCG ACGCTGACTG GACCCCACTG	CCCGGCGCAA	TCGGACCCCA GCGGACACCC	ATGCCCGATA	AAAGGACACA	ACATGGCTTT	CGTACATCAT	AATGGAAGCG	AGGGTGTCGC	GCCCCTGACC
	TGTACCGCCA	CATCCGGCGG	AGTCCTCGAT	ACGCTGACTG	ACGGCCAC ACCGTCACCG GACCCGCCGA CCCGGCGCAA		CCCCAACCGG	CCAGAACC GCACTTACCA ACCCTGCTTG AAAGGACACA	ACTGCGGCGA CGAATCGCCG ACATGGCTTT	GACGCCGTGC	GCCGACAACA	GTGCCCGTTC CTAATCGCCG AGGGTGTCGC	CGATCAACCG
FIGURE 3-8	CCGGGCGGAC	ATCACCCCGG	TCGTTGATCA	TGGTTCATCG	ACCGTCACCG	TTCCGGGCGG	CACCCCGACG	GCACTTACCA	ACTGCGGCGA	ACGCTGTGGT	CATGGGTGTC	GTGCCCGTTC	CCAGTGACCC
H	ACGGCCGCCA	CGCTGTCGCG	TGCGATCGTG			CAAAGCGTTG	CAAAGTGCTG	CGCCAGAACC	ACCACGCCGA	CGACCTGCAA	GCGAACCCGT		TAGAGGTCGA
	CTGGCCGTTC	CACCGGCGGT GCCGCATGAC CG	GTCCGATTCG CCTACGACTC TG	GCCCGCTCCT GGTCCGCGCA CA	CTGGCCGCCG AGCTGCGCTA CC	CAGCAGTGCA CCGACTGGGC CA	GCCGTGTACA GGGCTTTATC CAAAGTGCTG CACCCCGACG CCCCAACCGG ATGCCCGATA	CTGCAACAGC AGCTCAATGC CG	AGCCATGGCT GAAACCCCCG AC	CAACGCCGAT GTCGGTATGG CGACCTGCAA ACGCTGTGGT	CCTGCCGAAC CTGCAGACCG GCGAACCCGT CATGGGTGTC GCCGACAACA AATGGAAGCG	CGCGAACTGT CCCGTCGACG TCGGTAAGCC	CGACAGTACC GACGACACCA TAGAGGTCGA
	CGGTACCTAC	CACCGGCGGT	GTCCGATTCG	GCCCGCTCCT	CIGGCCGCCG	CAGCAGTGCA	GCCGTGTACA	CTGCAACAGC	AGCCATGGCT	CAACGCCGAT	CCTGCCGAAC	CGCGAACTGT	CGACAGTACC

FIGURE 3-9	3-9	Page 9 of 16	16
AACGACGAAC GACAACTGAT GCACGAGCTG GCAG	GCAGTCCAGG TTGTCTGCTC GCAGACGGGT		6300
TGCTCACCCG ATGCGGCGGT CGAAGCACTC GAAT	GAATCCTTCG CGAAAGACGG	AACACTTATC 63	6360
CTCCGCGGCG ACACCGAGAA CGCCTACCTC GAAG	CGCCTACCTC GAAGCCGGAG GCAATGTTCT TGTCCATGCC		6420
GATCGTGACT GGCTTGCCTT CCACGCGTCG TATC	CCACGCGTCG TATCCCGGCA ACGACCCGCT GCGAGACGCC		6480
CGACCTATCG AGCAGGACGA CGACCAGGGG GCGG	GCGGGGTCGC CATCGTGACC	AGGCCCAGCC 65	6540
CGGACACCGC CACGGTGCCG GCGCGCATGC ACGCT	ACGCTCATTA CCTAGACTAA	AAATTGATGG 66	0099
GAGGACCGAT GCCAAGACCA CCGAAACCGG CCCGC	CCGAAACCGG CCCGGCTCAA ACTGGTTGAG GGCCGCTCCC		0999
CCGGCCGCGA TTCCGGCGGC CGGAAAGTCC CCGA(CCGAGTCGCC GAAGTTTATC CGTCAGGCAC		6720
CGGATGCCCC GGACTGGCTC GACGCCGAGG CGCTC	CGCTGGCCGA ATGGCGGCGC	GTCGCACCGA 67	6780
CTTTGGAGCG GCTTGACCTG CTCAAACCTG AGGAI	AGGATCGGGC GCTCCTGTCC GCGTACTGCG		6840
AGACCTGGTC CGTCTACGTC GCGGCGGTTC AGCGC	AGCGGGTCCG CGCCGAAGGC	CTCACAATTA 69	0069
CCTCACCGAA ATCCGGTGTC GTGCACCGGA ACCCC	ACCCGGCGGT GACGGTTGCG	GAGACGCCGC 69	0969
GCATGCATCT GCTGCGCTTG GCCTCCGAGT TTGGC	TTGGCCTGAC CCCGGCCGCC GAGCAGCGAC		7020

Page
Pa
9
3-1
KE
JUE
—

			FIGURE 3-10		Page	10 of 16
TGGCGGTGGC GCCGGGCGAC	GCCGGGCGAC	GACGGCGACG	GGCTCAACCC	GACGGCGACG GGCTCAACCC GTTTGCCCCCG GACCGGTGAT	GACCGGTGAT	7080
GACCTTTTGT GTGTGATACA	GTGTGATACA	ATCGAGTTTG	GCATCTCGGC	ATCGAGTTTG GCATCTCGGC ATCCGCTGAC GCCGGGCAGT	GCCGGGCAGT	7140
CGCCGCGGGG CGGCTGGAAC	CGGCTGGAAC	CCGGATAGCG	GCCGCCATGC	CCGGATAGCG GCCGCCATGC GCCACAAGCG ATTCCGCGCG	ATTCCGCGCG	7200
TTTCTTGCGT CTGCTAGGTG	CTGCTAGGTG	GTGGCCGAAT	TTTGAGTAGC	TTTGAGTAGC ATCCTTTTCC GCATGGCCGA	GCATGGCCGA	7260
GCTGCGGTCT GGCGAAGGCC	GGCGAAGGCC	GAACCGTGCA	CGGCACCATC	GTGCCCTACA	ACGAGGCGAC	7320
CACCGTCCGC GACTTCGACG	GACTTCGACG	GCGAGTTCCA	GGAAATGTTC	GCGAGTTCCA GGAAATGTTC GCTCCTGGCG CTTTTCGGCG	CTTTTCGGCG	7380
CTCCATCGCC GAGCGCGGCC	GAGCGCGGCC	ACAAATTGAA	GCTGCTGGTC	ACAAATTGAA GCTGCTGGTC TCTCACGACG CTCGAACCCG	CTCGAACCCG	7440
CTACCCGGTG GGCCGGGCCG	9009990099	TTGAGTTGCG	GGAGGAGCCT	CACGGCTTGT	TCGGGGCGTT	7500
CGAGATTGCG GACACCCCGG	GACACCCCGG	ACGGCGACGA	GGCTTTGGCG	AACGTAAAAG	CTGGTGTCGT	7560
CGACTCGTTT TCGGTGGGTT	TCGGTGGGTT	TCCGACCGAT	CCGGGACCGT	TCCGACCGAT CCGGGACCGT CGCGAAGGGG ATGTGCTGGT	ATGTGCTGGT	7620
GCGCGTCGAA GCGGCGCTGT	GCGGCGCTGT	TAGAGGTTTC	CCTAACCGGC	CCTAACCGGC GTTCCGGCCT ATTCGGGGGC	ATTCGGGGGC	7680
ACAAATCGCC GGGGTGCGCG	GGGGTGCGCG	CGGAATCGCT	TACAGTCGTT	TCCCGTTCGA CAGCCGAAGC	CAGCCGAAGC	7740
CTGGCTGTCC CTACTCGATT	CTACTCGATT	GGTGAACAAT	CTATGACCGA	ATTCGACGAC	ATCAAAAACC	7800

		FI	FIGURE 3-11		Page	11 of 16	
TCTCTTTACC	TGAAACCCGT	GACGCGGCGA	AGCAGCTCCT	TCTCTTTACC TGAAACCCGT GACGCGGCGA AGCAGCTCCT CGACAGTGTC GCCGTGTGAC	GCCGTGTGAC	7860	
CTGACCGGTG	CTGACCGGTG AGGCGCGCA GCGTTATTCA	GCGTTATTCA		GGCGCTGACG CGCCACGCCG AGGAACTGCG	AGGAACTGCG	7920	
GGCGGAGCAG	GGCGGAGCAG CGCCGCGCG GCCGCGAAGC	GCCGCGAAGC	CGAGGAGGAG	CTGCGCCGCT	Acceeeccee	7980	
TGAGCTGAGG	GTGGTGCCCG	GCGCTCCCAC	CGGCGGCGAC	TGAGCTGAGG GTGGTGCCCG GCGCTCCCAC CGGCGGCGAC GACGGCGACG CGCCGCCGGG	ອອອນນອນນອນ	8040	
CAACTCGTTG	CGGGACACCG	CGTTTCGCAC	ACTGGATTCT	CAACTCGTTG CGGGACACCG CGTTTCGCAC ACTGGATTCT TGTGTGCGAG ACGGCCTGAT	ACGCCTGAT	8100	
GTCGTCGCGG	GCGCCGGAGA	CCGCGGAAAC	GTCGTCGCGG GCGGCGGAGA CCGCGGAAAC CTTGTGCCGC	ACCGGGCCGC CGCAGTCCAC	CGCAGTCCAC	8160	53/6
crcereeece	CTCGTGGGCG CAGCGCTGGC TGGCGGCCAC	TGGCGGCCAC	CGGCAGCCGC	GACTATTTGG	GCGCGTTCGT	8220	3
CAAGCGGGTT	TCCAATCCTG	TTGCGGGGCA	CACGGTTTGG	CAAGCGGGTT TCCAATCCTG TTGCGGGGCA CACGGTTTGG ACCGACCGGG AAGCGGCCGC	AAGCGGCCGC	8280	
GTGGCGTGAG	GCTGCCGCGG	TGGCCGCCGA	GCAGCGAGCG	GTGGCGTGAG GCTGCCGCGG TGGCCGCCGA GCAGCGAGCG ATGGGCCCTGG TGGACACCCA	TGGACACCCA	8340	
AGGCGGGTTT	AGGCGGGTTT CTGATCCCGG	CGGCGCTGGA	CCCGGCGATC	CTGCTGTCGG	GTGATGGGTC	8400	
GACGAACCCG	ATTCGGCAGG	TGGCGAGGGT	GACGAACCCG ATTCGGCAGG TGGCGAGGGT GGTGCAAACG ACCTCCGAGA	ACCTCCGAGA	TTTGGCGGGG	8460	
CGTGACTTCC	CGTGACTTCC GAAGGCGCCG AAGCTCGTTG	AAGCTCGTTG	GTACTCCGAA	GTACTCCGAA GCCCAGGAGG TGTCCGACGA	TGTCCGACGA	8520	
TTCGCCAGCG	TTGGCCCAGC CGGCGGTGCC			GAACTACCGT GGAAGCTGCT GGATTCCGTT	GGATTCCGTT	8580	

					54/6	3						
8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
AGATTCTCGC	ACGCCCAGCC	GCGCGGGGTC	CAAGGTTCCA	AGGCGGAAAC	TAGCCGGGAA	CGACGAATCA	GGTCCATGGT	GCGGATTCTT	TTCTGAAGGT	AGGGGTGCCG	NNNNNNNNN	NNNNNNNNN
GAGATCGGCA	GGCTCCGGCA	GTGGTCGTCG	SCGCTGCCGC	ACGTTGCGGC	CCGCCGATGC	GCGGTGACAG	GACAGAGTTG	ACCGGGCAGC	GCGTTTCGAG	GCCTCTGCTT	CGATTCGTNN	NNNNNNNNNN
CTTCGTTGGC	GTTCGTCAAC	CTCCGATCAG	GTTGCAGTCG		GCACGACAGT	CGTTGATTCG	CCTCATCGTC			GGCGTGGGCG	GTTGCGTTGT	NNNNNNNNNN NNNNNNNNNN NNNNNNNNN
		TAACCGGCAC	ATGTTTACGC	CGAACTTGTC	TCCCATCGCT	ACATGGACAC	Ö	₽	CAGATGTGCT	GATAGGGCCA	TCCCCTGCGG	NNNNNNNNNN
; CTGGAGGGTG	GAGCAACTGC	GICAGCGCGC	GTGGCGGCGG	GCGTTCGCGG	GCGCTGAAAT	GAAGTCTCCC	CTTGGCGACT	CCTCACCTGT	AGGGTCGGAT	GCGTAGGTAG	GCCCGCCAAC	NNNNNNNNN NNNNNNNNN NNNNNNNNN
CTCCATCGAG	GGACAGCGTT	CACCGGGTTC	AGAAGCGATT	GGCCAGCGCC	TTCGAATGGC	GTCTGTCCTG	TCCACTGGTG	GGAGTTGGTG	CGCCTGGTTC	GGAGACTACC	GGCCGGCCAC	NNNNNNNNN
	CTCCATCGAG CTGGAGGGTG ACGCGGCGAG CTTCGTTGGC GAGATCGGCA AGATTCTCGC 8640	ACGCGGCGAG CTTCGTTGGC GAGATCGGCA AGATTCTCGC AGACCGCGGC GTTCGTCAAC GGCTCCGGCA ACGGCGAGCC	ACGCGCCGAG CTTCGTTGGC GAGATCGGCA AGATTCTCGC AGACCGCGGC GTTCGTCAAC GGCTCCGGCA ACGGCGAGCC TAACCGGCAC CTCCGATCAG GTGGTCGTCG GCGCGGGGTC			8640 8760 8820 8880 8940	CTTCGTTGGCGAGATCGGCAAGATTCTCGC8640GTTCGTCAACGGCTCCGGCA8700CTCCGATCAGGTGGTCGTCG8760GTTGCAGTCGCAGCGGGGTC8820CACCATCAACACGTTGCGCAGGCGGAAACGCACGACAGTTAGCCGGAAAC8940GCACGACAGTCGCCGATGCTAGCCGGAACGTTGATTCGGCGGTGACAGCGACGAATCA	8640 8700 8760 8880 8940 9000	8640 8700 8760 8880 8940 9000	8640 8700 8760 8880 8940 9000 9060	8640 8700 8760 8880 8940 9000 9060 9120 9180	ACCCGCCGAG CTTCGTTGGC GAGATCGGCA AGATTCTCGC GACCGCGCG GTTCGTCAAC GGCTCCGGCA ACGGCGGGCC AACCGGCAC CTCCGATCAG GTGGTCGTCG GCGCGGGGTC TGTTTACGC GTTGCAGTCG GCGCTGCCGC CAAGGTTCCA GGAACTTGTC CACCATCAAC ACGTTGCGGC AGGCGGAAAC GCATGGACAC CGTTGATTCG GCGCTGCCGC AGGCGGAAAC CCCATCGCT GCACGACAGT CCGCCCGATGC TAGCCGGGAA CCCATCGCT GCACGACAGT CCGCCCGATGC GGCGGAATCA GAACTTGTC CACCACAGAGTTCG GCGTTGAACGT GAAGCCAATT CCTCATCGTC GACAGAGTTG GGTCCATGGT GGAGCCCGAA TCGCCGGCCG ACCGGGCAGC GCGGATTCTT AGATGTGCT GGTGCCCAAC GCGTTTCGAAGGT AGATGTGCT GGTGCCCAAC GCGTTTCGAAGGT AGATGTGCT GGTGCCCAAC GCGTTTCGAAGGT AGATGTGCT GGTGCGCAAC GCGTTTCGTNN NNNNNNNNNNNNNNNNNNNNNNNNNN

						55/6	3						
3 of 16	9420	9480	9540	0096	0996	9720	9780	9840	0066	0966	10020	10080	10140
Page 13	NNNNNNNNN	NNNNNNNNN	CCTGTTGCCG	CATTTGTTTC	TCGTTTGAGC	AGTGTGGCGC	ACCGAGGCGA	CCGGCGACCA	TGGGTTGCGT	GCGCCGATAG	GGCCCCAAAC	TTGGCTTTCG	AGATAGCTTC
	NINNINNINNIN NINNINNINNIN NINNINNINNIN NINNIN	NNNNNNNNN NNNNNNNNNN	NNNNNNNCC AAGCCAGAAT ATCGAGCCTG GCGGCCATGG TCGCCGCCTT	CTGCTTGGCT TTCGGCCGTT CCAGCTCGGC GATCCGGCGG CCAGCGGCGC CATTTGTTTC	CGTTGTCGAT	CCTCCTTGTC	TGATCGATGC ACCGAGGCGA	CCACCTCGC GTTGGCGCTC CTTGGCTTTC GGGCGTTCCA GCTCGGCGAT CCGGCGACCA	GTCGTCGCTG	GCCGAGGGCT	GTAGCCCATG GGCCCCAAAC	CTGCCGCTGC TTGGCTTTCG	TGTTTCTCCG
FIGURE 3-13	NNNNNNNNN		GCGGCCATGG	GATCCGGCGG	CTGCGGGTTG	GCTGCGCCTA	CCGGCATTGG	GGGCGTTCCA	GGATTTCTTT	CGGCGGAGAT	ATTCTGGTTT	ATGATCCT GCCCTCGCG	CGGCGCCATT
FIC	NNNNNNNNN	NNNNNNNNN	ATCGAGCCTG	CCAGCTCGGC	TTTGTCGTCG	GATGCCGAGG	GGGACCAAAG	CTTGGCTTTC	GCGAACCGGC	CGGTAGGTGC	ACGCTCTG		CGGCGGCCAG
	NNNNNNNNN	NNNNNNNNN NNNNNNNNNN NNNNNNNNN	AAGCCAGAAT	TTCGGCCGTT	TCCGCGAACA GGCGGATTTC TTTGTCGTCG	CGCTTGTAGG TGCCGGCGGA GATGCCGAGG	TGAGACGGCT TTGGTTCCAT GGGACCAAAG CCGGCATTGG	GTTGGCGCTC	GCGGCGCCAT TTGTTTCTCC GCGAACCGGC	TGTCGATTCG TTTGAGCCGC CG	CAGTGTCTGT TTTCGTCGAA TG	CAGAATATCG AGCCTGGCGG CC	CTCCGCGATC
	NNNNNNNNN	NNNNNNNNN	NNNNNNNCC	CTGCTTGGCT	TCCGCGAACA	CGCTTGTAGG	TGAGACGGCT	CCACCCTCGC	GCGGCGCCAT	TGTCGATTCG	CAGTGTCTGT	CAGAATATCG	GCCGCTCCAG CTCCGCGATC

						56/63	3						
14 of 16	10200	10260	10320	10380	10440	10500	10560	10620	10680	10740	10800	10860	10920
Page	GTCCTGTTGC	CCCGATCATC	TCCGAGTTGC	GCCGCACTCG	GAGCTCGCCG	CACCTCCAAG	GCAGACGGCG	TGGTGTGGTG	CCCCGCACCA	CACGACGATC	CGTCCTCGCC	GCGATCAAAC	CACGCGTGTT
	ACCGCCCAGC	CGCCGCCAAG	CGGCTCCACG GCGGGGTGTC GTCGCCGTCG GGCTCGTCGT CGCCGGCGAG TCCGAGTTGC	CGGCCCGGAG CGGCCGGCCA GGATCGCCGG GCCGCACTCG	CTGATCCGCA GGAACGCTTC GAGCTCGCCG	GCGCTTCCCT	GGCGCGTCCG GGAACCACGC CCGGATCGTC TCGGCCACCT GGTCGCGGTC GCAGACGGCG	TGAAACGTGC	CTTCAGCAGT CCACGGCCAA	CGAGGACAAC		CTGGCCATTA AGACGAAGTT	AGGTACTCGG
FIGURE 3-14	ATCCATGCCC CGCCCGTGGG	GCTGGGCTTC	GGCTCGTCGT	CGGCCGGCCA		AGCCGAACGA	TCGGCCACCT	ACCAAGTGGT	CTTCAGCAGT	GCCACAACGA	GACATCCCCA		TCACCCCAGC AGGTACTCGG
FI	ATCCATGCCC	CTCGTCTTCC	GTCGCCGTCG	CGGCCCGGAG	GTGCTCGACG	CGCGGCCTGC	CCGGATCGTC	GGTTACCGTC	GTGGAATGTT	AAGCTGCTAC	AAGATCTTTG	AACCCGCGAG	LGGGCCCGCC
	CGGCCCATGG GCCGGAAGCT	CGCGGTGTTC ACCGTCAGCG	GCGGGGTGTC	CGGTTGATGG CGGCTTGTTC	TCGCCGCCGG CGGCGTCGGC GTGCTCGACG	GTGTGCTCGT GCCGATTCAA CGCGGCCTGC	GGAACCACGC	CGGTCACCGG TITCCAGGTC GGTTACCGTC ACCAAGTGGT TGAAACGTGC	GTCATGGTTG ATCTCCTGGC GTGGAATGTT	ACACCTTCCA CCACCACGAG AAGCTGCTAC	CGTGAGAATC GCCGCCCGCG AAGATCTTTG GACATCCCCA CATCGACGTG	ACCTGGCCAG CACCCGCCG AACCCGCGAG	CCCTTCGCCA TCAAGCTTTT TGGGCCCGCC
	CGGCCCATGG	CGCGGTGTTC	CGGCTCCACG	CGGTTGATGG	DOCCECCE	GTGTGCTCGT	GGCGCGTCCG	CGGTCACCGG	GTCATGGTTG	ACACCTTCCA	CGTGAGAATC	ACCTGGCCAG	CCCTTCGCCA

					5	67/63							
15 of 16	10980	11040	11100	11160	11220	11280	11340	11400	11460	11520	11580	11640	11700
FIGURE 3-15 Page	ACCGTCAATC CGCGACGGCC GTCCCGCAAT	GACCCGTCGG GCCAGGGTGA CATGGGCGGT	CGGGCGCC AGGTGCGGGC CGCAGAGCCG	TGGTCGGC ACCACCAGCC GGGTGAACAC	GCCGATCACG CAGTCCAGCG GCAGCCGACG	CGGGGCGATC CGTTCGGCCA CCGCCAGCGA	GTATGCCG GCGGCGGCCA ACCCCGCCCA	CGAAGACC AGCTCGATCG AATGCACCAT	TCGAACGCCG CCGCGCTCAT CGCCGCGAAG	GGCAGAGCGG CCCGCACCGT AGCAATGCGC	CCCGGGCG GGTCCGGCCA GCTGCCGATC	AAGCGCTT CCAACGTCAA CTTGGTGTGG	ACCAAAGCCG CGGCGGCCAC GTCGACGGCG
FIGU	GCCGTCCCAG CGGCGCAAGC CGGCGAACCG AC	GCGCAGCGCC CGCCCCAATT GGTGACCACC GA	CCACTGACCG GGCAGGCTGT TGGCCATCGG CGCGGCGCC AGGTGCGGGC CGCAGAGCCG	GIGCACCICG GCAIGCAGGG CCAAAAGCIC GCIGGICGGC ACCACCAGCC GGGIGAACAC	GACATTGGCC CGCCCGAACA GCACCGGCGC GC	GGCAACCGCA CCCAGCGGCT CATCGACCTC CC	CACGTGCGGA CGGCTGGCCG GCGCCTGGCT GGGTATGCCG GCGGCGGCCA ACCCCGCCCA	GATGCGCCGG ATCGCCGCCT CGGTATCGCT GTCGAAGACC AGCTCGATCG AATGCACCAT	CAGCCGACCA GCCCGGCAAC CCAGTTGCGG TC	TCCCCGGCAT CCAGCGACGC GGCCCCGGCG GG	GCCAGCGCGG ACCGATTCGA GGCTGCCACC AACCCGGGCG GGTCCGGCCA GCTGCCGATC	ACCAGCCCTG CACATGAAAC CTGTTGTGCA GCAAGCGCTT	TTGAGGGTGC CCAGGTCCGC GGTGACCACC AC

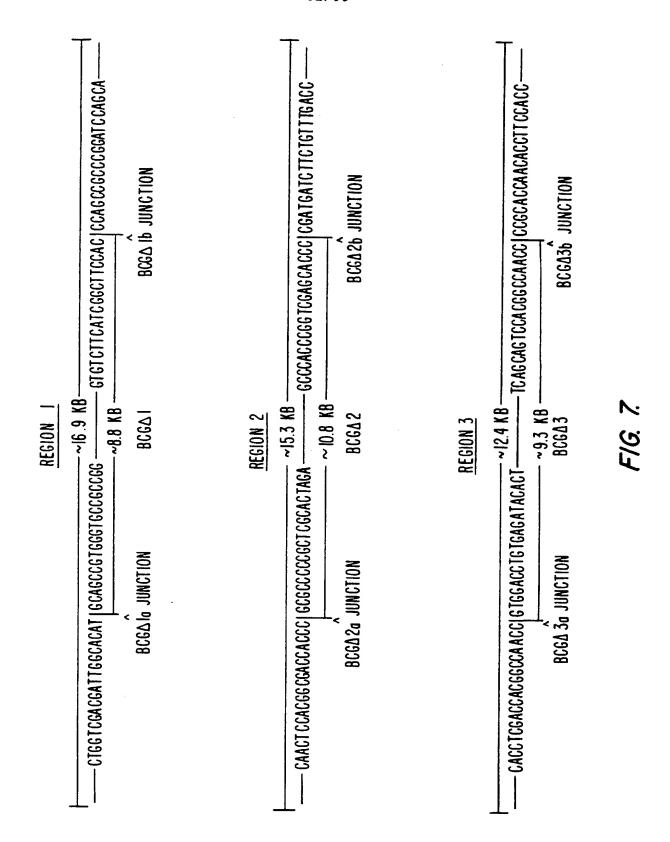
						58/	63					
16 of 16	11760	11820	11880	11940	12000	12060	12120	12180	12240	12300	12360	12412
FIGURE 3-16 Page 16	ACATCGCGCA GCGTGACGCC CGGCTCGGCG AGTTCGACCA GCAGCCCGCC CGCCCCCTCG	ACGAGGGTCA ACCGCCCGGG ACGGTCCAGG TCTGCGATCA GCCGCACGAT CTGATCGCGG	GCGGGCAACG CCATCCCGGC GTGTTCGGCG GCGCGGCCG GGGCCATCGG CTGCGGATAT	CGCGCCAAGC CGGCCAGCTG GGTCACCCCG GCCAACCGGC CGACCTCGGC GAGGTCGTCG	TCACCGCGGG CGGTGCCGGT CTGAACGGGC TTGCACACCG CCACGTCGAT GCCGGCCTGA	CGTGCGGCCG ACGCCAGCGC CGCGCAGACG ACCGTCTTGC CGACCCCCGT GCCGGTCCCG	GTGACGACCA GGATCGTCAA CGGCGCGCCA CGGCGAGAAC ATCCGTCAGC ACCCGCCGGG	CCAGCTCGAG CTCGCCGGCG TTCAGCGATG CGCGCGCGGT CAGCCGCAGC CGCGACGTAC	CCGCGGGCAC CGTCGGCGGC CGGAAGCAGC CCACCTTGAC CCCGGCGTCC AGGCAGGCCG	CCGCGGCGGC CACTGCCGAC TCCGGCTCGC CCAGGATCAC CGACACCATC GCCGAGTCCG	GCACCGCAGC CACACCGCAC ATCCGCGCAA GTTCACCAGC GTGGTTGAGC ACCGCCTGCG	ATCGCCACGG CTCGGCCTGC AAGACGCGCCA GCGCGGCCCG TGCGGCACCT TC
	ACATCGCGCA	ACGAGGGTCA	GCGGGCAACG	CGCGCCAAGC	TCACCGCGGG	CGTGCGGCCG	GTGACGACCA	CCAGCTCGAG	CCGCGGGCAC	ລອອລອອລອລລ	GCACCGCAGC	ATCGCCACGG

59/63 BCG∆ = ~ 8.8kb HOMOLOGUE ACCESSION 246257 U01072 X79562 129506 ORF 2.3e-43 1.4e-14 3.0e-13 3.6e-16 ≅ 9 HOMOLOGIES TO PREDICTED FOR ENCODED PROTEIN 三 B. subtilis subtilisin SERINE PROTEASES M. tuberculosis esat6 M. leproe aceA BCG uraA £ £ MAX.~kDa) ENCODED Protein 57 36 59 46 20 34 **BINDING SITES** ൊ NONE GGA(11) AGGAGA (10) POSSIBLE RIBOSOME AGGA (10) GAGG (5) 66A (4) 66A(9) A66A(9) GAA (5) 66A(5) NONE =14823 - 13438 14643 - 13438 14541 - 13438 3130-4203 3139-4203 5075-6046 夂 13328-11946 10619 - 9663 16190 - 14820 889-2433 6954-8612 2 1542 1368 £ 696 1386 102 1657 954 1380 **9** 鞷 ORF | M. tuberculosis | CODON USAGE REGION 1 (16.9 kb) YES YES \leq YES YES YES Æ YES YES 图 \cong ೨ 9 ೪ $\underline{\mathbf{H}}$ 王

F/G. 4.

							60/									
F/G 5.		BCG \(\times = \cdot \text{10.8 kb} \)	HOMOLOGUE Accession #				P24194	000015 A00975 U03393		X73226 X17445	A30545	X65104 U04851 Z22594 X59155				
≅-1	_ ક	908	P VALUE				9.9e -47 < le-5	1.5e-7 ~4e-5		9.9e - 146 2.7e - 36	6.7e - 141	3.1e - 11 1.4e - 08 4.4e - 11 2.5e - 09				
Bg Kp Xb	11 12 13 14 1	26 2H 21	HOMOLOGIES TO PREDICTED Encoded protein				E.col. iciA IysR FAMILY	MLEPRAE COSMID BIG20 ORF CUTINASES		Ecolisia Drouvwx	M. tuberculosis mpt 64	Ecol, gabP PERNEASE Styphimurium asp PERMEASE 7. harzianum indal gene RETROVIRAL RECEPTOR				
Bm Kp	02 6	35	ENCODED PROTEIN (MAX.~kOa)	25	91	34	34	22	61	37	24	2	31	35	25	12
Bm Sp Bm		20 2J 2E	POSSIBLE Ribosome Binding Sites	AGGGAG (7)	AGAA (4)	NONE	AG? (8) GGA (8)	NONE	NONE	AGGA (11)	AAGA (6)	AG (10)	GGAAGA (6)	GAG (10) GGAA (8) GGA (9)	NONE	AG (10)
H3 Bm	4 5 6	2L 2C 2K	START - STOP (BASE PAIRS)	1829-2386	2862-3298	3003-3590	5187-6134 5376-6134	6561-7217	8036-8560	9941-10909	11118-11783	11965-13407	14221-13376	8259-7211 7939-7211 7931-7211	4992-4327	5117-4521
.	2 3	2A 282 281	ORF SIZE (BASE PAIRS)	558	437	588	948	657	522	996	999	1443	846	1050	999	597
REGION 2 (15.3 KB) Bm Xb Bg			M. tuberculosis CODON USAGE	YES	YES	YES	YES	YES	YES		YES	YES	YES	YES	YES	YES
꿆			ORF	2A	281	282	32	82	2E	2F	23	24	2	2.1	2K	21.

				61/6	3										_
6	9.3 KB	HONOLOGUE ACCESSION #		090361 M29040 K00676 X01805, X07724		X76288	X76288 L 37531				010000	000010	M29292	0100010	
F16. 6.	BCGA = ~9.3 KB ORF	P VALUE	2.9e-64 5.1 e-13	3.0e-05 7.8e-4 8.2e-4		4.2e -26	3.2e - 11				6.2e-69	6.9e-53	1.4e-05	1.0e-81	
Ps Bg H3 R5 Xh	31	HONOLOGIES TO PREDICTED ENCODED PROTEIN	MTB mce sau3A M. leprae cosmid L247	ACTINOPHAGE R4 011P GENE A subfit site spec. Recomb. Recombinases/invertases		<i>S. coelicolor</i> phage phi-G31 EARLY REGION	paly palz GENES				M. leprae B1170	M. leprae bioDAYB	B. sphaericus bio DAYB	M. leprae cosmid BII70	
Ps CI)H	ENCODED PROTEIN (MAXkDa)	45	49	61	34	21	83	61	47	28	70	000	25	23
	31 36	POSSIBLE RIBOSOME BINDING SITES	NONE		66A (6)	NONE GGA (7)	NONE GGAA(5)	(A (9)	GAAGG (8)	GA (8)	GGAAG (6)	1117 440	GAA (II)	AGG (4)	AG (10)
Kp Rl	30 3E	START -STOP (BASE PAIRS)	613-1755	1214-2560	2820 - 3332	4007 - 4930 4070 - 4930	4795-5337 4915-5337	5639-6214	5253-7762 7285-7762	7868 - 9197	10146 - 11810 10164 - 11810	0.00	7019-11402	11594-10893	10147-179488
2	38	ORF SIZE (BASE PAIRS)	1143	1347	513	924	543	576	510	1330	9991	e e	8) 5)	702	099
REGION 3 (12.4 KB) RI Kp 1	3A	M. tuberculosis CODON USAGE	YES	YES	YES	22	NO VES	YES	YES	YES	YES		조 <u>.</u>	YES	YES
		ORF	3A	38	35	8	. H.	7.	36	3	31		₹;	3	3



SUBSTITUTE SHEET (RULE 26)

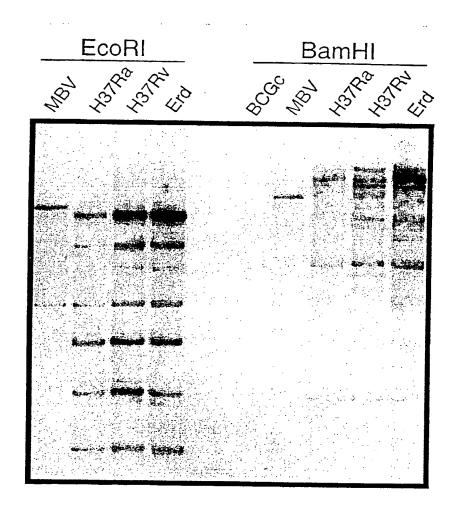


Figure 8

International application No. PCT/US96/01938

			101/0590/019	
	ASSIFICATION OF SUBJECT MATTER			
	:Please See Extra Sheet. :Please See Extra Sheet.			
	to International Patent Classification (IPC) or to both	national classification	and IPC	
	LDS SEARCHED			
Minimum c	documentation searched (classification system followe	d by classification sym	bois)	
U.S. :	435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350			24.32, 24.33
Documenta	tion searched other than minimum documentation to th	e extent that such docum	nents are included	in the fields searched
			· · · · · · · · · · · · · · · · · · ·	in the fields scarcing
Electronic o	data base consulted during the international search (na	ame of data base and, v	where practicable	search terms used)
	ee Extra Sheet.	,		, 1111111111111111111111111111111111111
		· · · · · · · · · · · · · · · · · · ·		
	CUMENTS CONSIDERED TO BE RELEVANT			
Category*	Citation of document, with indication, where as	opropriate, of the releva	ant passages	Relevant to claim No.
X	Infection and Immunity, Volume	61, No. 5, is	sued May	1-10, 16, 17,
	1993, H. Li et al, "Evidence for ab	sence of the MI	PB64 gene	24, 25
Υ	in some substrains of Mycobacto	erium bovis BC	G", pages	
	1730-1734, see entire document.			18-23
X	JP, 1-247094 (AJINOMOTO ET A	U) 02 October	1989 500	1-7
	entire document.	ic, oz octobei	1303, 366	1-7
x	Infection and Immunity, Values 5	0 N 40 '		
^	Infection and Immunity, Volume 5 1991, C. Parra et al, "Isolation	9, NO. 10, ISSUE	ed October	1-17
	molecular cloning of a specific m	on, characteriz	ation and	
	antigen gene: identification of a sp	ycobacterium to	perculosis	
	pages 3411-3417, see entire doci	iment	, gongrou	
	page street out, bod silling dock	Sincin.		
X Furth	ner documents are listed in the continuation of Box C	. See patent	family annex.	
A dox	ecial categories of cited documents: cument defining the general state of the art which is not considered	date and not in c	oublished after the inte onflict with the applica ory underlying the inve	rnational filing date or priority ation but cited to understand the
	be of particular relevance lier document published on or after the international filing date			claimed invention cannot be
·L· dox	current which may throw doubts on priority claim(s) or which is	considered nove	l or cannot be conside tent is tuken alone	red to involve an inventive step
Cite	ed to establish the publication date of another citation or other ecial reason (as specified)	"Y" document of pa	rticular relevance; the	e claimed invention cannot be
	cument referring to an oral disclosure; use, exhibition or other ans	considered to i	involve an inventive	step when the document is a document, such combination
	cument published prior to the international filing date but later than priority date claimed		per of the same patent	
Date of the	actual completion of the international search	Date of mailing of the	international sea	rch report
17 APRIL	. 1996	29 MAY 1	996	
	nailing address of the ISA/US	Authorized officer	Minal	Frus 10
Box PCT	ner of Patents and Trademarks	JEFFREY FRED		7/000 / / /
Washington Facsimile N	n. D.C. 20231 o. (703) 305-3230			
	. (.5), 555 5250	Telephone No. (70)3) 308-0196	

International application No. PCT/US96/01938

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
Category	Chanon of document, with indication, where appropriate, of the relevant passages	Relevant to claim 140
Y	Abstracts of the 1994 IDSA Annual Meeting, Clin. Infect. Dis., Volume 19, issued October 1994, R. Frothingham et al, "Sequence based strain differentiation in the Mycobacterium tuberculosis complex, including rapid identification of M. bovis BCG", page 565, see abstract 10.	1-25
X	R. GHERNA et al, "AMERICAN TYPE CULTURE COLLECTION: CATALOGUE OF BACTERIA AND PHAGES", Eighteenth edition, published 1992, pages 202 and 211, see entire document.	11-15
X	Infection and Immunity, Volume 62, No. 4, issued April 1994, L.	1-7, 16-25
	Pascopella et al, "Use of in vivo complementation in	
Y	Mycobacterium tuberculosis to identify a genomic fragment associated with virulence", pages 1313-1319, see entire document.	26
Y	Science, Volume 261, issued 10 September 1993, S. Arruda et al, "Cloning of an M. Tuberculosis DNA fragment associated with entry and survival inside cells", pages 1454-1457, see entire document.	1-23
X	US,A,5,171,839 (PATARROYO) 15 December 1992, columns 5-	1-10
	10.	16.22
Y		16-23
Y	Nature, Volume 256, issued 07 August 1975, C. Kohler et al, "Continuous cultures of fused cells secreting antibody of predefined specificity", pages 495-497, see entire document.	10
Y	US,A, 4,683,202 (MULLIS) 28 July 1987, see entire document.	16-22, 24, 25
Y	Genomics, Volume 4, issued 1989, D. Wu et al, "The ligation amplification reaction (LAR) amplification of specific DNA sequences using sequential rounds of template directed ligation", pages 560-569, see figure 2.	16-22, 24, 25
Y	US,A, 4,410,660 (STRAUS) 18 October 1983, columns 14 and 15.	23
Y	Gene, Volume 131, issued 1993, A. Kinger et al, "Identification and cloning of genes differentially expressed in the virulent strain of mycobacterium tuberculosis", pages 113-117, see page 114, column 2.	1-26
X,P	WO,A2,95/17511 (JACOBS ET AL) 29 June 1995, see entire	1-26
- - , -	document	

International application No. PCT/US96/01938

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
(,E	J. Bacteriol., Volume 178, No. 5, issued March 1996, G. Mahairas et al, "Molecular analysis of genetic differences between mycobacterium bovis BCG and virulent M. bovis", pages 1274-1282, see entire document.	1-26
′, P	Microbiology, Volume 141, issued 1995, J. Rodriguez et al, "Species-specific identification of mycobacterium bovis by PCR", pages 2131-2138, see entire document.	1-7, 16-22, 24, 25
X	Hybridoma, Volume 13, No. 1, issued 1994, A. Arya et al,	8-10
··	"Production and characterization of new murine monoclonal antibodies reactive to mycobacterium tuberculosis", pages 21-30, see page 27, table 1.	16-23

International application No. PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C12Q 1/68; G01N 33/53; C12P 19/34; C12N 5/10, 1/21; C07K 5/00, 14/00, 16/00; C07H 21/02, 21/04

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.3, 24.32, 24.33

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE, BIOSIS, CAPLUS, WPIDS

search terms: mycobacter?, tubercul?, bovis?, BCG, calmette, guerin, DNA, RNA, oligo, nucleic, oligonucleotide, hybridi?, probe, primer, amplif?, PCR, polymerase chain, ligase chain, LCR, attenuat?, immunoassay, antibod?, monoclon?, polyclon?, protein, peptide, antigen, virulenc?, infect?

Form PCT/ISA/210 (extra sheet)(July 1992)*

Value of Proce

```
standard; DNA; PRO; 1069 BP.
IĎ
     AF004671
XΧ
     AF004671;
4G
XX
     q3253155
NI
XX
     29-JUN-1998 (Rel. 56, Created)
ЭT
     29-JUN-1998 (Rel. 56, Last updated, Version 1)
TC
XX
     Mycobacterium tuberculosis H37Rv esat6 promoter region, L45 antigen
DE
     homologous protein LHP (lhp) gene, complete cds, and early secreted antigenic target 6 kDa (esat6) gene, partial cds.
DΕ
DΕ
XX
KW
XX
     Mycobacterium tuberculosis
OS
     Eubacteria; Firmicutes; Actinomycetes; Mycobacteria; Mycobacteriaceae;
OC
     Mycobacterium.
\circC
XX
     [1]
RN
     1-1069
RP
     Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;
RA
     "Promoter analysis of the M. tuberculosis orf1C gene encoding the early
RT
     secreted antigenic target 6 kDa (ESAT-6)";
RT
     Unpublished.
RL
XX
RN
      [2]
     1-1069
RP
     Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;
RA
RT
     Submitted (19-MAY-1997) to the EMBL/GenBank/DDBJ databases.
RL
     Mycobacterial Genetics Unit, Institut Pasteur, 25, rue de Dr Roux,
RL
     Paris, 75 75724, France
RL
XX
                       Location/Qualifiers
FH
      Key
FH
                       1. .1069
FT
      source
                       /organism="Mycobacterium tuberculosis"
FT
                       /chromosome="Region of difference RD1"
FT
                       /strain="H37Rv"
FT
                       525. .827
FT
      CDS
                       /codon_start=1
FT
                       /db_xref="PID:g3253156"
FT
                       /note="culture filtrate protein 10 kDa CFP-10"
 FT
                       /transl_table=11
 FT
                       /gene="lhp"
 FT
                       /product="L45 antigen homologous protein LHP"
 FT
                       /translation="MAEMKTDAATLAQEAGNFERISGDLKTQIDQVESTAGSLQGQWR
 FT
                       AAGTAAQAAVVRFQEAANKQKQELDEISTNIRQAGVQYSRADEEQQQALSSQMGF"
 FT
                       860. .>1069
 FT
      CDS
                       /codon_start=1
 FT
                       /db_xref="PID:g3253157"
 FT
                       /note="secreted T cell antigen; ESAT-6"
 FT
                       /transl_table=11
 FT
                       /gene="esat6"
 FT
```

BESTFIT of: Sa202820_0001.Dna check: 1179 from: 1 to: 1277 (i) APPLICANTS: (A) NAME: INSTITUT PASTEUR (A) NAME: STATENS SERUM INSTITUT (ii) TITLE OF INVENTION: A POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS from: 1 to: 1069 check: 4782 to: R55u027.Af0046**71** AF004671 standard; DNA; PRO; 1069 BP. ID XXAF004671; ACXX q3253155 NIXX Symbol comparison table: Gencoredisk: [Gcgcore.Data.Rundata] Swgapdna.Cmp CompCheck: 2335 Average Match: 10.000 50 Gap Weight: Average Mismatch: -9.000 3 Length Weight: 1069 Length: 10595 Quality: Gaps: Ratio: 9.911 Percent Identity: 99.532 Percent Similarity: 99.532 Match display thresholds for the alignment(s): = IDENTITY 5 1 Sa202820_0001.Dna x R55u027.Af004671 October 21, 1998 14:36 ... 1 CTGCAGCAGGTGACGTCGTTGTTCAGCCAGGTGGGCGGCACCGGCGGCGG 50 1 CTGCAGCAGGTGACGTCGTTGTTCAGCCAGGTGGGCGCCACCGGCGGCGG 51 CAACCCAGCCGACGAAGCCGCGCAGATGGGCCTGCTCGGCACCAGTC 100 51 CAACCCAĠĊĊĠĀĊĠĀĠĠĀĀĠĊĊĠĊĠĊĀĠĀŤĠĠĠĊĊŤĠĊŤĊĠĠĊĀĊĊĀĠŤĊ 100 151 GGCCTGCTGCGCGGAGTCGCTACCTGGCGCAGGTGGGTCGTTGACCCG 200

201 CACGCCGCTGATGTCTCAGCTGATCGAAAAGCCGGTTGCCCCCTCGGTGA 250
251 TGCCGGCGGCTGTTGCCGGATCGTCGGTGACGGGTGGCGCCGCTCCGGTG 300
301 GGTCCGGGAGCGATGGGCCAGGGTTCGCAATCCGGCGGCTCCACCAGGCC 350
351 GGGTCTGGTCGCGCCGCCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
401 AGGACGACTGGGACGAAGAGGACGACTGGTGAGCTCCCGTAATGACAACA 450
451 GACTTCCCGGCCACCCGGGCCGGAAGACTTGCCAACATTTTGGCGAGGAA 500
501 GGTAAAGAGAAAGTAGTCCAGCATGGCAGAGATGAAGACCGATGCCGC 550
551 TACCCTCGGGCAGGAGGCAGGTAATTTCGAGCGGATCTCCGGCGACCTGA 600
601 AAACCCAGATCGACCAGGTGGAGTCGACGGCAGGTTCGTTGCAGGGCCAG 650
651 TGGCGCGCGGCGGGGCCGCCCAGGCCGCGGTGGTGCGCTTCCA 700
651 TGGCGCGCGGCGGGGACGGCCGCCCAGGCCGCGGTGGTGCGCTTCCA 700 701 AGAAGCAGCCAATAAGCAGAAGCAGGAACTCGACGAGATCTCGACGAATA 750
701 AGAAGCAATAAGCAGAAGCAGGAACTCGACGACGAGTCTCGACGAATA 750 701 AGAAGCAGCCAATAAGCAGAAGCAGGAACTCGACGAGATCTCGACGAATA 750
751 TTCGTCAGGCCGGCGTCCAATACTCGAGGGCCGACGAGGAGCAGCAGCAG 800
801 GCGCTGTCCTCGCAAATGGGCTTCTGACCCGCTAATACGAAAAGAAACGG 850

851	AGCAAAAACATGACAGAGCAGTGGAATTTCGCGGGTATCC	900
		900
901	GGCAAGCGCAATCCAGGGAAATGTCACGTCCATTCATTCCCTCCTTGACG	950
901	GGCAAGCGCAATCCAGGGAAATGTCACGTCCATTCATTCCCTCCTTGACG	950
951	AGGGGAAGCAGTCCCTGACCAAGCTCGCAGCGGCCTGGGGCCGGTAGCGGT	1000
951		
- 001	TCGGAGGCGTACCAGGGTGTCCAGCAAAAATGGGACGCCACGGCTACCGA	1050
1001		
1001	TCGGAGGCGTACCAGGGTGTCCAGCAAAAATGGGACGCCACGGCTACCGA	1050
1051	GCTGAACAACGCGCTGCAG 1069	
1051	GCTGAACAACGCGCTGCAG 1069	

.